



Contrôle de la croissance et régulation génique chez *Escherichia coli*

Jerome Izard

► To cite this version:

Jerome Izard. Contrôle de la croissance et régulation génique chez *Escherichia coli*. Sciences agricoles. Université de Grenoble, 2012. Français. NNT : 2012GRENV061 . tel-01124077

HAL Id: tel-01124077

<https://theses.hal.science/tel-01124077>

Submitted on 6 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Virologie, Microbiologie, Immunologie**

Arrêté ministériel : 7 août 2006

Présentée par

Jérôme Izard

Thèse dirigée par **Hans Geiselmann**

et codirigée par **Stephan Lacour / Delphine Ropers**

préparée au sein du **Laboratoire Adaptation et Pathogénie des Micro-organismes**

et de l'**Ecole Doctorale Chimie et Sciences du Vivant**

Contrôle de la croissance et régulation génique chez *Escherichia coli*

Thèse soutenue publiquement le **6 décembre 2012**,
devant le jury composé de :

Mr, Ariel Lindner

Chargé de recherche, INSERM, Président

Mr, Steve Busby

Professeur, University of Birmingham, Rapporteur

Mme, Muriel Cocaign-Bousquet

Directeur de recherche, INRA, Rapporteur

Mr, Hans Geiselmann

Professeur universitaire, UJF, Directeur de thèse

Mr, Stephan lacour

Maître de conférence universitaire, UJF, Co-Directeur de thèse

Mme, Delphine Ropers

Chargé de recherche, INRIA, Co-Directeur de thèse



Remerciements

Tout d'abord, je tiens à remercier le Docteur Muriel Cocaïgn-Bousquet et le Professeur Steve Busby pour m'avoir fait l'honneur d'être les rapporteurs de mon manuscrit de doctorat. Je remercie également le Docteur Ariel Lindner d'avoir accepté de faire partie de mon jury.

J'exprime ma gratitude au Docteur Marie-France Cesbron-Delauw et au Professeur Hans Geiselmann pour m'avoir accueilli au laboratoire durant ces 4 dernières années.

Je fais part de ma reconnaissance à mes trois directeurs de thèse : Delphine Ropers, Stephan Lacour et Hans Geiselmann pour leur aide, leur disponibilité, les conseils qu'ils m'ont prodigués et leur soutien tout au long de ma thèse. Je tiens également à les remercier pour la formation scientifique de qualité qu'ils m'ont apportée.

Un grand merci à toute l'équipe TaMaRa, pour m'avoir reçu à Paris pendant 3 semaines. Un merci tout particulier à Yifan qui a pris beaucoup de son temps pour me former à l'utilisation de leur microscope et pour m'avoir initié à la microfluidique.

Je remercie sincèrement toutes les personnes du LAPM avec qui j'ai pu échanger et travailler tout au long de ma thèse. Notamment, Elisabeth, Joël, Mickael, Sébastien, Thomas et Xavier qui m'ont fait profiter de leur expérience dans le monde la recherche.

Je souhaite aussi remercier chaleureusement les membres de l'équipe CEG : Corinne, Caroline, Claire, Julien, Yves et Woei-Fuh avec qui j'ai partagé des moments exceptionnels autant au niveau scientifique que humain. Je remercie également les différents stagiaires qui ont apporté leur dynamisme et leur énergie : Laure, Clément et Mathias. Ainsi que les personnes sans qui le laboratoire ne fonctionnerait pas aussi bien : Christiane, Evelyne et Severine.

J'adresse aussi ma gratitude aux autres membres de l'équipe IBIS, en particulier pour leurs encouragements et leur soutien lors de la rédaction de ce manuscrit.

Je remercie mes collègues et désormais amis : Colin, Maxime, Omayya et Guillaume pour nos interminables discussions au laboratoire ou au O'Callaghan, leur énergie et leur

infaillible soutien.

Je ne serais jamais parvenu à cette étape majeure de mon cheminement personnel et professionnel, sans le support de ma famille et de mes proches. Je tiens à remercier de tout cœur ma maman et ma sœur « Doudou » pour m'avoir soutenu et poussé tout au long de cette grande aventure. Enfin je remercie Laura, ma compagne, pour sa patience et son soutien jusqu'aux derniers instants, je n'y serais pas arrivé sans toi.

Table des matières

Remerciements	iii
Table des matières	v
Table des figures	vii
Liste des tableaux	ix
Introduction	3
1 Croissance bactérienne	3
1.1 Conditions de croissance en laboratoire	3
1.1.1 Nutriments et conditions de culture	3
1.1.2 Modes de culture	6
1.1.2.1 Culture en milieu non renouvelé	6
1.1.2.2 Culture continue	7
1.2 Dynamique de la croissance	9
1.2.1 Phases de croissance	9
1.2.2 Mesure de la croissance	12
1.3 Adaptation à l'environnement	13
2 Régulation de l'expression génique	17
2.1 Régulation globale	19
2.1.1 ARN polymérase	19
2.1.1.1 Régulations de la production de l'enzyme cœur	19
2.1.1.2 Spécificité des facteurs sigmas	21
2.1.1.3 Compétition entre facteurs σ	22
2.1.1.4 Modulation de l'affinité du facteur σ^{70}	23
2.1.1.5 Crl : modulateur de l'activité de σ^{38}	24
2.1.2 Ribosomes	24
2.1.2.1 Organisation et régulation des opérons ribosomiaux	25
2.1.2.2 Réponse stringente et rôle de ppGpp	27
2.1.2.3 Assemblage et cycle de traduction	29

2.1.2.4	Hibernation des ribosomes en phase stationnaire	31
2.1.3	Distribution de l'ARN polymérase et des ribosomes	33
2.2	Régulation spécifique de l'expression génique	34
2.2.1	Régulation par des facteurs de transcription	34
2.2.2	Régulation par les protéines associées au nucléoïde (NAPS)	37
2.2.2.1	Modes de répression par les NAPs	38
2.2.2.2	Modes d'activation par les NAPs	40
	Objectifs	41
	Résultats	45
3	Étude de l'expression de régulateurs globaux : cas du gène <i>crl</i>	45
3.1	Mesure dynamique de l'expression des gènes	45
3.2	Analyse des cinétiques d'expression	47
	Publication n°1 : « WellReader : a MATLAB program for the analysis of fluo- rescence and luminescence reporter gene data »	51
3.3	Identification de nouveaux régulateurs transcriptionnels	61
3.4	Étude de la régulation du gène <i>crl</i>	61
	Publication n°2 : « CRP and Fis regulate the expression of the Crl regulatory protein in <i>Escherichia coli</i> »	65
4	Effet de la machinerie globale d'expression sur la croissance	95
4.1	Rôle de la machinerie globale sur l'expression génique	95
4.2	Contrôle du taux de croissance en modulant la machinerie globale	97
4.2.1	Modulation de l'activité traductionnelle	97
4.2.2	Modulation de l'activité de l'ARN polymérase	100
	Publication n°3 : « Growth control of <i>Escherichia coli</i> by the global transcrip- tional machinery »	107
	Discussion et perspectives	131
	Annexe	135
	Publication n°4 : « A genome-wide screen for identifying all regulators of a target gene »	137
	Bibliographie	168

Table des figures

1.1	Effet de la température sur la croissance de différentes classes d'organismes . .	5
1.2	Dispositif de micro-fluidique permettant d'observer la croissance à long terme d' <i>E. coli</i>	8
1.3	Phases de croissance d'une culture en batch	9
2.1	Effet du taux de croissance sur les paramètres cellulaires globaux et sur l'ex- pression de gènes constitutifs.	18
2.2	Structure génique de l'opéron <i>rpoBC</i> chez <i>E. coli</i>	20
2.3	Organisation des opérons d'ARN ribosomiaux chez <i>E. coli</i>	26
2.4	Structure générale de l'opéron <i>rrnB</i> d' <i>E. coli</i>	27
2.5	Redistribution de l'ARN polymérase en fonction des signaux environnementaux	28
2.6	Métabolisme du (p)ppGpp chez <i>Escherichia coli</i>	30
2.7	Cycle d'hibernation des ribosomes au cours de la croissance.	32
2.8	Complexité du réseau de régulation transcriptionnelle d' <i>Escherichia coli</i>	36
2.9	Exemples de répression et d'activation par les protéines du nucléoïde.	39
3.1	Plasmide rapporteur contenant le promoteur du gène <i>crl</i> en amont de l'opéron luciférase	46
3.2	Mesure dynamique de l'expression génique en microplaque	47
3.3	Exemple de traitement de données obtenues avec la fusion <i>crl</i> : :lux	50
4.1	Modèle mécanistique représentant la machinerie globale d' <i>E. coli</i>	96
4.2	Effet de l'expression des protéines Rmf, Hpf et YfiA sur l'activité traductionnelle	99
4.3	Constructions et plasmides utilisés pour contrôler le niveau d'ARN polymérase	101
4.4	Contrôle de la croissance en induisant l'expression de <i>rpoBC</i> et échappement des cellules	103

Liste des tableaux

1.1	Composition macromoléculaire d' <i>Escherichia coli</i> lors d'une croissance exponentielle à 37°C	14
2.1	Affinités des différents facteurs sigma pour le cœur de l'ARN polymérase chez <i>E. coli</i>	22

Introduction

Croissance bactérienne

Les bactéries se reproduisent par fission binaire au cours de laquelle une cellule mère se divise en deux cellules filles génétiquement identiques. Chez la bactérie *Escherichia coli*, on observe ainsi un allongement des cellules, sous forme de bâtonnets de 2 à 4 μm environ, taille à laquelle les cellules se divisent. Le temps requis pour permettre à une cellule de se diviser est très variable et dépend de nombreux paramètres génétiques et environnementaux. Dans le cas d'*E. coli*, ce temps peut varier d'une quinzaine de minutes à plusieurs heures.

1.1 Conditions de croissance en laboratoire

1.1.1 Nutriments et conditions de culture

Chaque organisme doit trouver dans son environnement les nutriments nécessaires pour croître et se diviser. En laboratoire, on peut faire croître de nombreuses espèces bactériennes dans des milieux de culture qui apportent en solution les nutriments essentiels à la croissance. Au niveau élémentaire, les besoins nutritifs d'une bactérie comme *E. coli* sont multiples et tous ne sont pas requis dans les mêmes quantités. Ainsi les éléments comme : C, H, O, N, P, S, K, Mg, Ca, Na, Fe doivent être présents en grandes quantités, tandis que d'autres seulement sous forme de traces : Zn, Co, Cu et Mo. Ces éléments se trouvent sous forme d'eau, d'ions organiques, de petites molécules et de macromolécules.

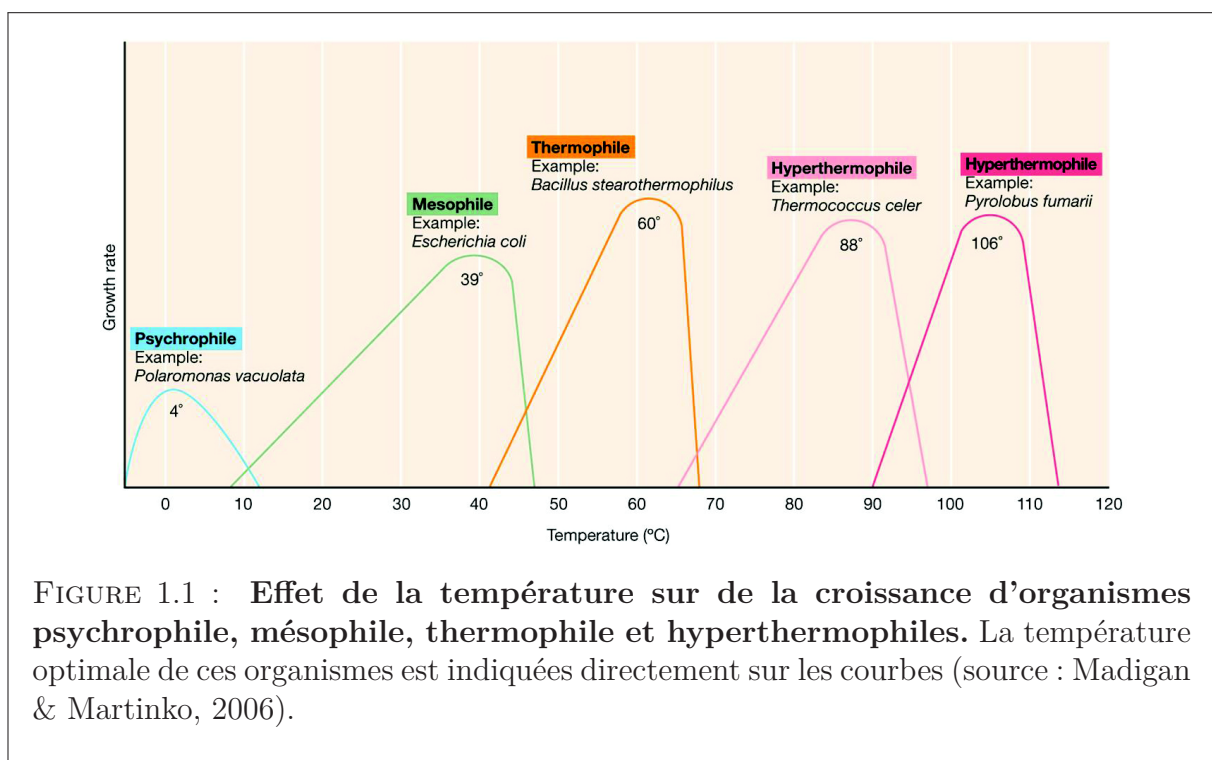
Parmi ces nutriments, les principaux sont le carbone et l'azote, qui représentent respectivement 50% et 12% de la masse sèche d'une cellule de microorganisme (Madigan & Martinko, 2006). Le phosphore est requis pour la biosynthèse des acides nucléiques et des phospholipides, le fer joue un rôle majeur dans la respiration cellulaire. Il est en effet un composant clef des cytochromes et protéines fer-soufre impliqués dans la chaîne de transport d'électrons. Ces nutriments servent à produire l'ensemble des composants de la cellule. On comprend alors que la composition du milieu de culture est un paramètre essentiel à la croissance.

En laboratoire, il existe deux grands types de milieux de culture : les milieux définis et complexes (ou non-définis). Pour un milieu défini, on connaît la composition chimique exacte de chacun des éléments qui le compose, à l'inverse d'un milieu complexe.

De nombreuses études sont réalisées dans un milieu complexe connu sous le nom de Luria-Bertani broth ou LB. Il a été formulé par Giuseppe Bertani en 1951, pour étudier la lysogénie chez *Escherichia coli*. Il l'appela alors « Lysogeny Broth » Bertani (1951). Par la suite plusieurs recettes en furent dérivées, notamment pour ne plus mettre de glucose comme c'était le cas au début. Le LB est un milieu très populaire du fait qu'il offre une croissance rapide des cellules, une division toute les 20 min pour la plupart des souches d'*Escherichia coli*. De plus sa composition est simple, il contient des produits de digestion de la caséine, plus communément appelés peptone ou tryptone, d'extrait de levure et de sodium chloride. Il est cependant déconseillé d'utiliser un milieu complexe pour caractériser la croissance ou la physiologie cellulaire. C'est pourquoi au cours de ma thèse, j'ai essentiellement utilisé du milieu minimum M9 (Sambrook & Russell, 2001). Il permet de mieux contrôler la croissance bactérienne en maîtrisant par exemple la concentration de glucose.

Après les nutriments, la température est l'un des facteurs environnementaux les plus importants pour la croissance et la survie des bactéries. Si la température est trop élevée ou à l'inverse trop basse, il leur sera impossible de croître. A mesure que la température augmente, les réactions chimiques et enzymatiques s'exécutent plus rapidement et la croissance cellulaire est plus rapide. Néanmoins, au dessus d'une certaine température, certaines protéines subissent des dommages irréversibles. Ainsi, il a été montré qu'un organisme possède une température minimale en dessous de laquelle il ne croît pas, une

température optimale, à laquelle la croissance est la plus rapide, et une température maximale au delà de laquelle la croissance est impossible. Ces trois températures sont appelées températures cardinales et sont propres à chaque type d'organisme. Cependant, elles ne sont pas complètement fixes et peuvent, par exemple, être altérées par la composition du milieu culture. Il est possible de classer les micro-organismes en fonction de leurs températures cardinales. On trouve alors quatre grandes classes selon leur températures optimales : les psychrophiles, avec une faible température, les mésophiles, avec une température intermédiaire, les thermophiles, avec une température élevée et les hyperthermophiles, avec une température extrêmement forte (Fig. 1.1). Les mésophiles sont généralement trouvés dans des animaux à sang chaud, et dans des environnements situés dans des zones tempérées et tropicales. Les hyperthermophiles sont retrouvées dans des habitats extrêmes tels que les sources chaudes, les geysers ou encore près des cheminées hydrothermales océaniques.



A l'instar de la température, chaque organisme possède une gamme de pH lui permettant de croître de façon optimale. La plupart des environnements ont un pH entre 5 et 9, et l'on retrouve souvent des organismes avec un pH optimum dans cette gamme. Quelques organismes, en particulier des champignons qui sont plus acido-tolérants que

les bactéries, présentent une croissance optimale à très faible pH et sont appelés acidophiles. On retrouve aussi des organismes vivant sous des conditions de pH très élevées, les alkaliphiles.

Lors d'une croissance en laboratoire, le pH du milieu est amené à varier au cours du temps. En effet, certains composants, acides ou alcalins, sont consommés et des produits issus du métabolisme sont également relâchés dans le milieu. C'est pourquoi dans les milieux de culture, il est courant d'ajouter un composé chimique, appelé « tampon », qui permet de conserver le pH relativement constant. Pour une gamme de pH proche du neutre (pH6-7,5), on utilise souvent du phosphate qui est un excellent tampon, généralement sous la forme KH_2PO_4 .

La concentration en oxygène présente dans le milieu est un autre paramètre important pour la croissance. Certains organismes ont besoin d'oxygène pour croître. Ces aérobies obligatoires utilisent cet oxygène pour produire de l'ATP par respiration aérobie. D'autres ne sont pas capable d'utiliser l'oxygène, et produisent de l'ATP par fermentation, ce sont des anaérobies. Il existe aussi des organismes capables de produire de l'ATP, soit par fermentation, soit par respiration, c'est le cas d'*Escherichia coli*, qui appartient au groupe des anaérobies facultatives. Habituellement les organismes aérobies ont une croissance plus rapide lorsque le milieu de culture est agité, ceci permet une meilleure oxygénation du milieu que par simple diffusion.

1.1.2 Modes de culture

Il existe plusieurs moyens de cultiver des bactéries, que l'on peut départager selon le type de renouvellement du milieu. En effet, on peut cultiver des cellules dans une condition où le milieu est soit renouvelé au cours du temps, on parle alors de culture continue, soit non renouvelé ou culture en « batch ».

1.1.2.1 Culture en milieu non renouvelé

Un moyen courant de cultiver des cellules est de les ensemercer dans une flasque, un tube ou encore un puits de micro-plaque contenant du milieu frais liquide. C'est ce qu'on appelle une culture en batch : le milieu est défini au départ et n'est pas renouvelé au cours du temps. Ceci implique que la composition chimique du milieu change, les nutriments

sont consommés au fur et à mesure par les bactéries, leur concentration diminue alors progressivement. De plus, des déchets métaboliques produits par les cellules s'accumulent dans le milieu, pouvant même stopper la croissance s'ils sont en trop grande concentration.

Avec ce mode de culture, il est possible de définir la qualité nutritionnelle du milieu, en prenant un milieu riche comme le LB ou un milieu minimum, dans lequel on pourra, par exemple, contrôler le type et la concentration de source carbonée. On peut également contrôler la température, en plaçant les cultures dans un environnement thermostaté, ainsi que l'oxygénation du milieu, en agitant les cultures. Couramment, on souhaite que ces deux paramètres demeurent constants tout au long de la croissance.

Il existe également une autre façon de cultiver des cellules dans un milieu renouvelé : sur boîte contenant un milieu gélosé (grâce à de l'agar). Ce milieu est utilisé la plupart du temps pour des cultures aérobies. En effet, on peut étaler des bactéries sur une boîte et lorsqu'elles se divisent, elles forment des colonies bactériennes visibles à l'œil nu.

Il est possible de faire varier la concentration d'agar, et l'on peut alors observer des phénotypes différents. Pour caractériser la motilité, par exemple, on utilise de faibles concentrations d'agar, les bactéries peuvent alors se déplacer à la surface de la boîte pour former des colonies très grandes avec un diamètre très important. Ce mode de culture est couramment utilisé pour isoler des cellules provenant d'une culture liquide. Néanmoins, il est peu utilisé pour caractériser la croissance du fait de son hétérogénéité, on lui préfère les cultures en milieu liquide.

1.1.2.2 Culture continue

Lorsque l'on cultive des bactéries en batch, l'inconvénient majeur est que la composition du milieu évolue au cours du temps. Le moyen le plus courant pour y remédier est de maintenir la culture dans un milieu continuellement renouvelé. On utilise pour cela un chemostat. Cet appareil est constitué d'un réservoir dans lequel se trouvent les cellules et le milieu, et pour lequel on contrôle l'arrivée, en ajoutant du milieu frais, et la sortie en retirant du milieu de culture et des cellules (Novick & Szilard, 1950). Ceci permet de contrôler indépendamment le taux de croissance et la densité cellulaire contrairement au mode de culture en batch où les deux sont fixés dès le départ (par la nature du milieu, le volume de la culture et sa dilution). A noter que si le taux de dilution est trop important,

le chemostat sera « lavé » de ses cellules, celles-ci se divisant à leur vitesse maximale – déterminée par la nature du milieu – mais ne suffisant pas à renouveler la culture. À l'inverse, si le taux de dilution est trop faible les cellules ne se diviseront plus car elles n'auront pas assez de nutriments. Un des avantages majeurs de la culture en chemostat est que l'utilisateur peut contrôler indépendamment la densité de la culture et sa vitesse de croissance, par exemple étudier une culture à forte densité présentant un faible taux de croissance ou encore maintenir une population en division active.

Plus récemment, grâce aux avancées faites dans le domaine de la microfluidique, l'étude du comportement des fluides en dessous du millimètre, il est possible de cultiver des cellules dans des micro-chemostats tout en les observant sous un microscope. Plusieurs exemples d'application peuvent être mentionnés, par exemple : l'étude des comportements bactériens face à des molécules chemo-attractantes (Ahmed *et al.*, 2010). Grâce à la capacité de créer des gradients extrêmement précis de ces molécules, la microfluidique devient un outil idéal pour l'étude de la motilité ou du chimiotactisme. Cette technique permet également d'étudier les mouvements individuels de bactéries (Sokolov *et al.*, 2010). Enfin, il est également possible d'étudier la croissance à long terme d'une population en observant chaque cellule de façon individuelle (Wang *et al.*, 2010 ; Fig. 1.2).

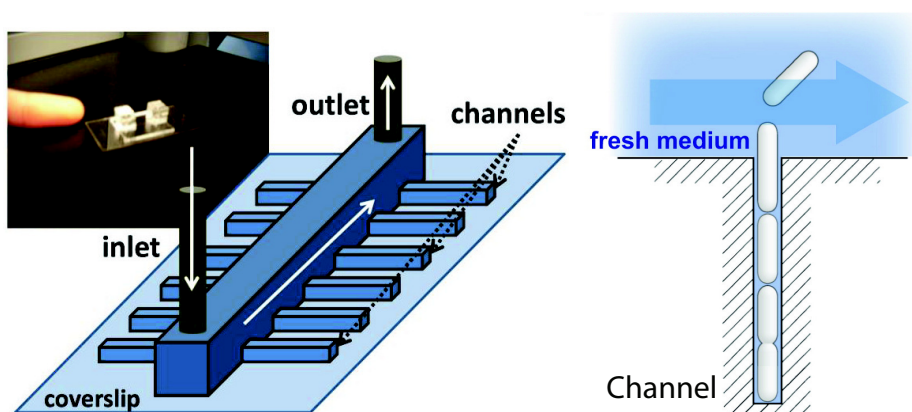
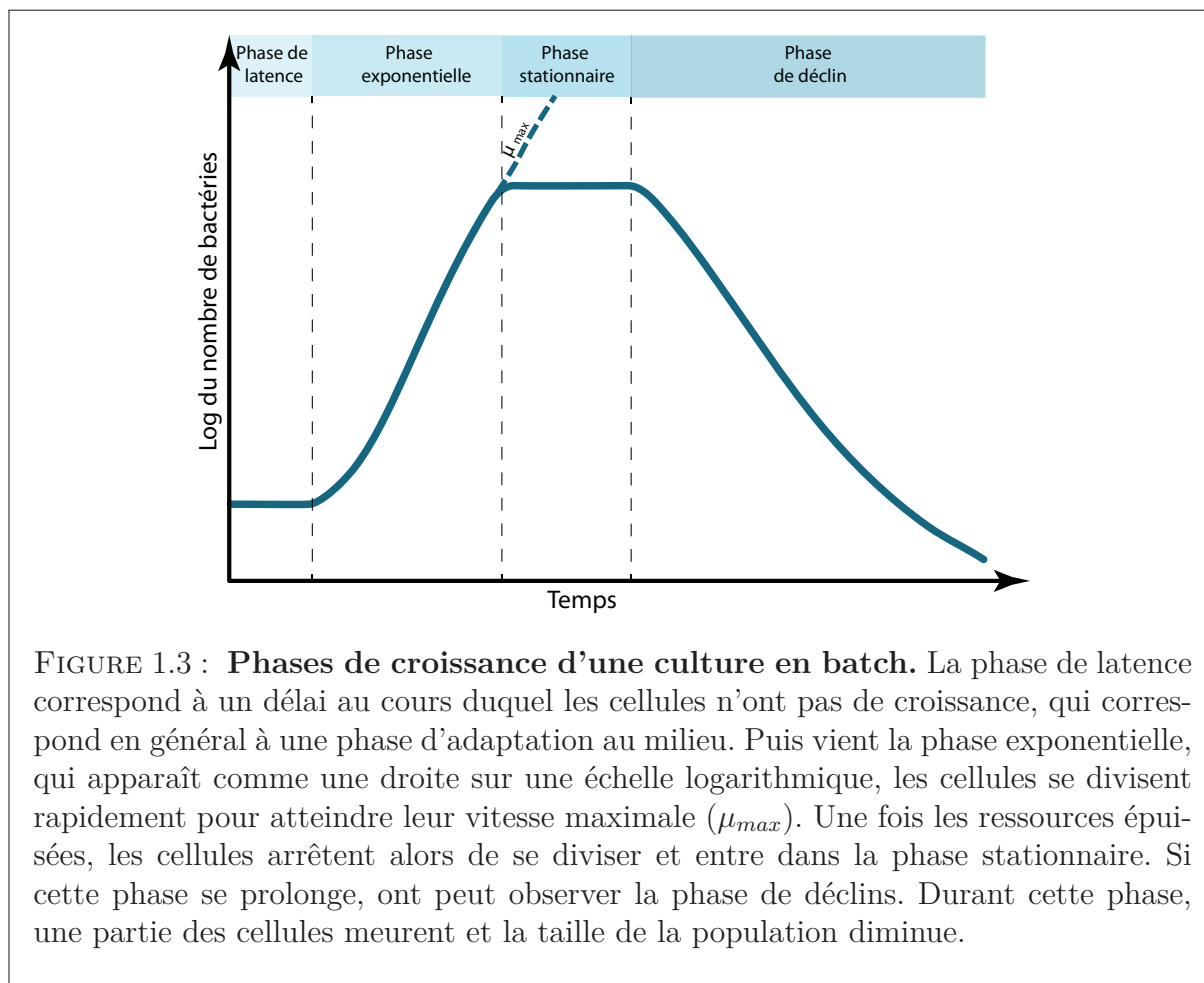


FIGURE 1.2 : Dispositif de micro-fluidique permettant d'observer la croissance à long terme d'*E. coli* (source : Wang *et al.*, 2010).

1.2 Dynamique de la croissance

1.2.1 Phases de croissance

Lors d'une culture en batch, nous avons vu que la composition du milieu évolue au cours du temps. Ainsi une succession de phases, différenciées par leur taux de croissance, peuvent être observées : la phase de latence, exponentielle, stationnaire et de déclin (Fig. 1.3).



Phase de latence. Lorsque l'on inocule une culture dans un milieu frais, la croissance ne démarre pas toujours immédiatement, mais souvent après un certain délai. Ce temps de latence dépend de l'état de l'inoculum et des conditions de culture. Si l'inoculum est constitué de cellules en phase stationnaire, alors il est fort probable d'observer une phase de latence importante avant la reprise de croissance. Plusieurs raisons peuvent expliquer ce phénomène : une fraction importante de cellules peut être non viable (Hinshelwood,

1946), les enzymes et autres constituants nécessaires pour croître dans le milieu frais peuvent être absents ou en quantité trop faible et doivent donc être synthétisés de nouveau. Aucune latence n'est observée lorsque l'on transfère des cellules en phase exponentielle de croissance dans un même milieu frais. On peut également observer une phase de latence, dans le cas d'une diauxie (croissance dans un milieu composé de deux sources carbonnées différentes, par exemple glucose-lactose). Elle correspond alors au temps nécessaire à la cellule pour synthétiser les transporteurs et enzymes spécifiques à l'utilisation du second sucre.

De même, on observe un délai avant la reprise de croissance lorsque l'on transfère des cellules d'un milieu vers un autre, par exemple d'un milieu riche (LB) vers un milieu minimum (M9 glucose). Le milieu étant plus pauvre, les cellules doivent produire par elles-mêmes les constituants essentiels qu'elles trouvaient auparavant dans le milieu (acides aminés, nucléotides etc.) (Monod, 1949). Le temps de synthétiser les enzymes permettant cette production provoque la latence.

Phase exponentielle. Durant cette phase, les cellules se divisent à leur taux maximal. Comme une cellule se divise pour en donner deux nouvelles, le nombre d'individus au sein de la population augmente de façon exponentielle. Une méthode courante pour visualiser et identifier cette croissance est de tracer le nombre d'individus sur une échelle semi-logarithmique en fonction du temps et si la courbe obtenue est une droite alors la population croît de façon exponentielle. Il est courant d'utiliser des cellules pour des études enzymatiques lorsqu'elles sont dans cette phase de croissance du fait que leur taux de croissance est constant et que les cellules sont donc probablement dans un état d'équilibre (steady state).

Durant cette phase, l'évolution de la population étant exponentielle, il est possible de calculer certains paramètres caractérisant la croissance. Du fait de la progression géométrique de la population, il y a une relation directe (expliquée par l'équation 1.1) entre le nombre de cellules présentes en début de phase exponentielle (N_0) et en fin de phase exponentielle (N) et le nombre de générations (n).

$$N = N_0 2^n \tag{1.1}$$

On peut alors simplement calculer le temps de génération g des cellules : $g = t/n$, où t

représente la durée de la phase exponentielle en unité de temps. A partir du temps de génération, il est possible de déterminer le taux de croissance μ en utilisant l'équation 1.2, qui correspond au nombre de générations par unité de temps pendant que des cellules sont en phase exponentielle.

$$\mu = \frac{\ln 2}{g} \quad (1.2)$$

A partir de ces paramètres, il est possible de caractériser la croissance d'une bactérie dans différentes conditions de culture, pourvu que les cellules soient en phase exponentielle. En particulier pour mesurer les effets positifs ou négatifs de traitement sur le croissance, comme par exemple la mutation d'un gène, l'ajout d'un composé dans le milieu.

Phase stationnaire. Lorsqu'un nutriment essentiel vient à manquer, les cellules arrêtent de se diviser : le taux de croissance devient nul et la taille de la population ne change plus, on dit alors que la population a atteint la phase stationnaire. On observe également l'entrée en phase stationnaire, lorsqu'un déchet ou une molécule produite par les cellules atteint une concentration inhibant la croissance. Cette phase n'apparaît donc que dans les modes de culture sans renouvellement du milieu. Le fait que les cellules ne se divisent plus ne veut pas dire qu'elles ne sont plus métaboliquement actives ou que l'ensemble des processus de biosynthèse est stoppé. Au contraire, elles mettent en place des stratégies d'adaptation à la carence et de résistance au stress, en modifiant leur métabolisme et leur expression génique. Des études ont même montré que certaines sous-populations de cellules apparaissent au cours de longue culture en phase stationnaire (Steinhaus & Birkeland, 1939). On note par exemple, l'apparition du phénotype GASP (Growth Adaptation in Stationary Phase) ; les cellules issues d'une longue phase stationnaire présentent des mutations qui leurs confèrent un avantage de croissance lorsqu'elles sont mises en compétition avec d'autres cellules. Ce phénotype a été découvert en mélangeant des cellules provenant d'une phase stationnaire courte et d'autre d'une phase longue de 10 jours (Zambrano *et al.*, 1993). On peut également observer dans certain cas un phénomène appelé « phase cryptique ». Lors de cette phase la taille de la population n'évolue plus mais une partie des individus meurt et se lyse, servant alors de substrat à une autre partie de la population qui continue de se diviser.

Phase de déclin. Cette phase de croissance est observée lorsqu’une culture reste longtemps en phase stationnaire. Les nutriments venant à manquer sont de plus en plus nombreux et le nombre de cellules qui meurent et se lysent devient très important et n’est plus compensé par les cellules qui continuent de se diviser. Le taux de croissance devient alors négatif.

1.2.2 Mesure de la croissance

Il existe plusieurs méthodes permettant de compter ou d’estimer le nombre d’individus d’une population bactérienne. Il est par exemple possible de compter directement le nombre de cellules d’un échantillon sous microscope ou de déterminer le nombre de cellules par unité de volume d’une culture en suspension en utilisant une lame de Malassez. C’est une méthode efficace pour estimer le nombre d’individus d’une population. Néanmoins, sans marquage spécifique, il est impossible de distinguer les cellules viables. Les cultures contenant peu de cellules doivent être concentrées puis observées afin de pouvoir les compter. Cette méthode est donc mal adaptée à des cultures de faible densité.

On peut également utiliser la cytométrie en flux pour compter directement le nombre de cellules d’une culture en suspension. Cette méthode est moins fastidieuse et permet de compter un plus grand nombre de cellules (échantillon plus grand). Elle présente toutefois le même inconvénient que la microscopie pour distinguer les cellules viables des autres cellules, il faut un marqueur fluorescent pour marquer les cellules viables ou mortes.

Une autre méthode consiste à étaler des cellules cultivées en milieu liquide, après plusieurs dilutions en cascades, sur un milieu gélose et à compter ensuite le nombre de colonies formées après incubation. Chaque colonie représente alors une seule cellule isolée provenant de la culture initiale. En sachant quelle dilution a été utilisée, ainsi que le volume ayant été étalé, on peut facilement remonter au nombre de cellules par volume de culture. Toutefois, cette méthode permet d’identifier seulement les cellules viables (et cultivables). Pour les cultures à fortes densités, il est nécessaire de faire plusieurs dilutions en cascades, ce qui engendre parfois des erreurs assez importantes. Enfin ceci ne concerne pas *Escherichia coli*, mais pour de nombreuses espèces bactériennes prélevées dans l’environnement (par exemple dans les sols ou dans le milieu marin) il est difficile de trouver des conditions permettant leur culture en laboratoire. Ainsi ces espèces sont viables mais

non cultivables et n'apparaissent donc pas dans le dénombrements des colonies.

Les méthodes décrites ici présentent l'avantage d'avoir une mesure directe du nombre de cellules d'une culture. Néanmoins, elles nécessitent pour certaines de gros appareillage ou encore sont longues et fastidieuses à mettre en place. Il existe donc des méthodes plus rapides, mais indirectes, permettant d'estimer le nombre de cellules d'une culture, par exemple en mesurant sa turbidité. Une culture cellulaire présente un aspect turbide car les cellules dispersent la lumière passant à travers la suspension. Plus l'échantillon est dense plus la lumière sera dispersée. Cette turbidité peut donc être mesurée à l'aide d'un spectrophotomètre, qui mesure la lumière non dispersée après être passée à travers un échantillon. On peut utiliser plusieurs longueurs d'onde pour mesurer la turbidité. Au cours de ma thèse j'ai mesuré l'absorbance à une longueur d'onde de 600 nm, on parle alors de densité optique (DO_{600}).

Pour une culture bactérienne, la densité optique est proportionnelle au volume de l'ensemble des cellules (Volkmer & Heinemann, 2011) et peut donc être utilisés pour déterminer le nombre de cellule d'un échantillon. Néanmoins, il est nécessaire de dresser pour chaque organisme une courbe standard reliant mesures directes (via microscopie ou comptage par exemple) et mesures de turbidité. Cette méthode présente l'avantage d'être rapide, relativement précise et de ne pas détruire ou modifier de façon importante les cultures. Elle est donc largement utilisée lorsque l'on veut mesurer l'évolution du nombre d'individu au cours du temps. Le même échantillon peut être lu à plusieurs intervalles de temps et reporté pour ensuite calculer un taux de croissance, par exemple.

1.3 Adaptation à l'environnement

Les bactéries présentent des taux de croissance très différents selon les milieux dans lesquels elles se trouvent. Néanmoins, il a été montré que la composition macromoléculaire des cellules, à une température donnée, dépend du taux de croissance et non de la composition du milieu. En effet, des cellules ayant des taux de croissance identique, cela même dans des milieux de culture différents, présentent la même composition macromoléculaire (Schaechter *et al.*, 1958 ; Maaloe, 1979). Ainsi les cellules contiennent plus d'ADN, d'ARN et de protéines lorsqu'elles poussent rapidement. Cela a été confirmé plus tard par Bremer

et Dennis, qui ont mesuré la composition cellulaire d'*Escherichia coli*, pendant la phase exponentielle de croissance à 37°C et dans différents milieux de culture (Tableau. 1.1). Ils ont ainsi pu obtenir des mesures à différents taux de croissance compris entre 0,6 et 2,5 doublements par heure.

TABLE 1.1: **Composition macromoléculaire d'*Escherichia coli* lors d'une croissance exponentielle à 37°C** (d'après Bremer & Dennis, 1996).

Paramètre	Unité	Taux de croissance μ (dbl/h)				
		0,6	1,0	1,5	2,0	2,5
Protéines	$\mu g/10^9$ cellules	100	156	234	340	450
ARN	$\mu g/10^9$ cellules	20	39	77	132	211
ADN	$\mu g/10^9$ cellules	7,6	9,0	11,3	14,4	18,3
Fourches de réplication	nb/cellule	1,46	2,14	3,64	5,92	9,19
Termini	nb/cellule	1,23	1,37	1,54	1,74	1,94
ARN polymérase	10^3 /cellule	1,5	2,8	5,0	8,0	11,4
Ribosome	10^3 /cellule	6,8	13,5	26,3	45,1	72,0

Il semble y avoir une dépendance plus forte entre la composition cellulaire et le taux de croissance qu'entre le type de milieu et la composition cellulaire. C'est pour cette raison que la plupart des paramètres de la cellule sont exprimés en fonction du taux de croissance (Bremer & Dennis, 1996). L'ajustement du contenu en ADN, ARN et protéine de la cellule est expliqué par des variations de la machinerie globale d'expression qui comprend l'ARN polymérase, pour la transcription, et les ribosomes, pour la traduction. On constate par exemple que la quantité d'ARN par cellule augmente de 10 fois lorsque l'on passe d'un taux de croissance de 0,6 à 2,5 doublements par heure, alors que le ratio ARN/ARN polymérase par cellule reste sensiblement le même (13 à 18) (Tableau 1.1; pour revue Bremer & Dennis, 1996). Il a été montré que des gènes dits constitutifs (ayant des promoteurs non régulés spécifiquement) présentent une expression dépendante du taux de croissance (Wanner *et al.*, 1977; Liang *et al.*, 1999).

Cet ajustement de la composition cellulaire, en fonction du taux de croissance, provient certainement des mécanismes de régulation de la machinerie cellulaire. Dans le chapitre suivant, nous nous intéresserons donc aux régulations globales de l'expression, comment

la synthèse et l'activité de l'ARN polymérase et des ribosomes sont régulées en fonction du taux de croissance. Puis nous aborderons les régulations plus spécifiques de l'expression des gènes, qui contribuent également à l'ajustement de l'expression génique en fonction des conditions environnementales.

Régulation de l'expression génique

Nous avons vu que la composition cellulaire varie avec le taux de croissance. Les composants qui dépendent du taux de croissance, comme par exemple le nombre de copies des gènes ou encore l'abondance des ribosomes et de l'ARN polymérase, affectent directement l'expression des gènes (Bremer & Dennis, 1996). En retour, les changements dans l'expression des gènes entraînent souvent des changements de taux de croissance. On comprend ainsi que croissance et régulations géniques sont intimement connectées.

Plusieurs études ont montré que des gènes constitutivement exprimés, c'est-à-dire dont les promoteurs n'ont aucune régulation spécifique connue (par exemple le promoteur PL du phage lambda ou celui du gène de résistance à l'ampiciline *bla*, présentent une expression dépendante du taux de croissance (Liang *et al.*, 1999 ; Wanner *et al.*, 1977). Cette relation de dépendance a même été expliquée de façon quantitative par un modèle ne faisant intervenir que des paramètres physiologiques généraux (Klumpp *et al.*, 2009). Ces paramètres comprennent les taux de transcription et traduction dans la cellule : le premier augmente lorsque le taux de croissance est grand, tandis que le second n'est pas affecté par la croissance (Fig. 2.1 A et B). Enfin le « gene dosage » c'est-à-dire le nombre de copie de gènes par cellule augmente avec le taux de croissance. Cela provient du fait que le nombre de fourches de réplication augmente plus la cellule se divise vite et par conséquent le nombre de copies des gènes proches de l'origine de réplication (Fig. 2.1 C). En utilisant ces paramètres généraux, Klumpp *et al.* ont par exemple montré que le nombre de transcrits d'un gène constitutif par cellule augmente proportionnellement

au taux de croissance (Fig. 2.1 D), tandis que le nombre de protéines par cellule reste constant quelque soit le taux de croissance (Fig. 2.1 E). Ceci montre que l'expression des gènes n'ayant pas de régulation spécifique est tout de même affectée par la croissance via la machinerie d'expression globale. Dans la section suivante, nous décrivons plus en détails la composition et la régulation des principaux composants de la machinerie cellulaire.

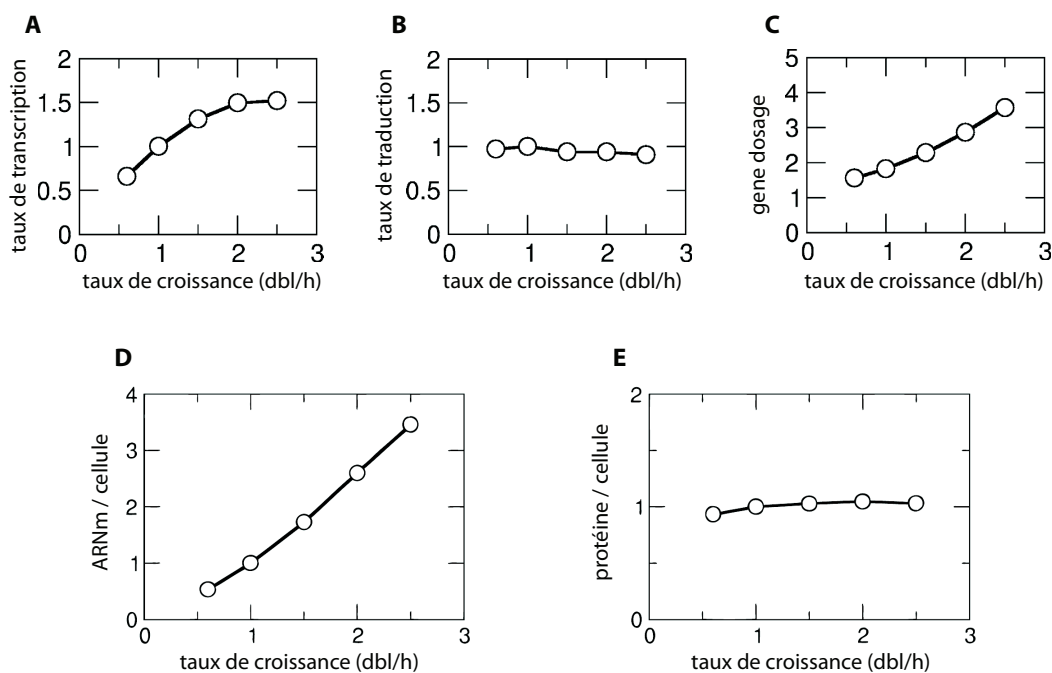


FIGURE 2.1 : Effet du taux de croissance sur les paramètres cellulaires globaux et sur l'expression de gènes constitutifs (source : Klumpp *et al.*, 2009). Taux de transcription par gène (A), taux de traduction par molécule d'ARNm (B) et nombre de copie de gènes par cellule (C) en fonction du taux de croissance. Tous ces paramètres ont été mesurés pour des gènes non-régulés, tels que le promoteur du gène de résistance à l'ampicilline *Pbla*, le promoteur PL du phage lambda, ou encore les promoteurs contrôlant la répllication du plasmide pBR322 (Liang *et al.*, 1999). Les graphes D et E montrent, respectivement, le nombre moyen d'ARN messagers et de protéines par cellule pour ces gènes en fonction du taux de croissance.

2.1 Régulation globale

2.1.1 ARN polymérase

Chez *E. coli*, il existe une seule ARN polymérase constituée d'une enzyme cœur contenant les sous-unités $\alpha 2\beta\beta'$ et ω . Cette enzyme cœur, avec une masse moléculaire d'environ 380kDa, catalyse la synthèse d'ARN à partir d'un fragment d'ADN. Pour reconnaître un promoteur et être capable d'initier la transcription, l'enzyme cœur fixe un des sept facteurs sigma existant chez *E. coli* pour former l'holoenzyme (Burgess, 1969 ; Gross *et al.*, 1992).

Trois étapes sont alors nécessaires pour réaliser un cycle de transcription complet : initiation, élongation et terminaison. Lors de l'initiation l'holoenzyme se fixe spécifiquement sur deux hexamères nucléotidiques, relativement conservés d'un promoteur à l'autre, situés en -35 et -10 par rapport au site initial de la transcription (+1) pour former un complexe fermé. La décaténation par le facteur σ ouvre le promoteur dans la région -10. Ce complexe ouvert, en présence de NTPs va commencer la synthèse d'ARN courts sans quitter le promoteur : c'est l'initiation abortive (Record *et al.*, 1996). Après plusieurs initiations abortives, le complexe va entrer en élongation, pour produire l'ARN messager. Au cours de cette phase, l'holoenzyme va perdre contact avec le promoteur et l'élongation se poursuit jusqu'à ce que le complexe rencontre un signal de terminaison, une séquence terminatrice d'ADN. Le complexe, l'ADN et l'ARN se dissocient alors de façon irréversible.

2.1.1.1 Régulations de la production de l'enzyme cœur

Pour constituer le cœur d'une ARN polymérase fonctionnelle, il faut que les différentes sous-unités s'assemblent selon une certaine stœchiométrie. Or la sous unité α semble être en excès de l'ordre de deux fois par rapport aux sous-unités β et β' (Engbaek *et al.*, 1976 ; Hayward & Fyfe, 1978 ; Iwakura *et al.*, 1974 ; Pedersen & Reeh, 1976). C'est pourquoi l'assemblage de l'enzyme cœur dépend essentiellement de la concentration des sous-unités β et β' .

Les gènes codant les sous-unités β et β' , respectivement *rpoB* et *rpoC*, se présentent sous la forme d'un opéron complexe contenant en amont 4 gènes codant les protéines



FIGURE 2.2 : **Structure génique de l'opéron *rpoBC* chez *E. coli*.** On comptabilise trois unités transcriptionnelles : la première qui contient les gènes *rplKA* (en vert), une deuxième contenant les gènes *rplJL* (en rouge), toutes deux codent des protéines ribosomales. Enfin une dernière, contenant les gènes *rpoB* et *rpoC* qui codent les sous-unités β et β' de l'ARN polymérase. En aval de *rpoC*, il y'a un terminateur de la transcription (boucle en trait plein). Un atténuateur de la transcription (boucle en trait pointillé) est également présent entre *rplL* et *rpoB*.

ribosomales RplK, A, J et L (Fig. 2.2). Dans ce cluster de gènes, on retrouve plusieurs promoteurs responsables de la transcription de *rpoBC*, le premier situé en amont de *rplK* et le second en amont de *rplJ* (Steward & Linn, 1991). De plus, un atténuateur transcriptionnel a été mis en évidence dans la région inter-cistronique *rplL-rpoB* (Barry *et al.*, 1979). Cette organisation complexe, avec plusieurs promoteurs, est propice à une régulation de la transcription de *rpoBC* (Steward & Linn, 1991 ; 1992), même si les mécanismes de régulation restent encore mal connus.

Plusieurs études ont mis en évidence que la synthèse de l'ARN polymérase est soumise à une auto-régulation (Yura & Ishihama, 1979). En particulier, l'holoenzyme ainsi que le complexe $\alpha 2\beta\beta'$ répriment l'expression des sous-unités β et β' dans un système de transcription-traduction *in vitro* (Fukuda *et al.*, 1978). Plus tard, d'autres résultats toujours *in vitro* ont suggéré que cette auto-régulation se fait au niveau traductionnel (Kajitani *et al.*, 1980 ; Peacock *et al.*, 1982). Enfin, il a également été montré *in vivo* que les sous-unités β et β' s'auto-régulent (Bedwell & Nomura, 1986 ; Dennis *et al.*, 1985 ; Meek & Hayward, 1986 ; Passador & Linn, 1989). Par exemple, la traduction de l'ARN messager *rpoBC* est augmentée entre 2 et 3 fois lorsque la synthèse de β et β' est limitée (Dennis *et al.*, 1985). À l'inverse, l'augmentation de la synthèse d'ARN polymérase, en exprimant l'ensemble des sous-unités depuis un plasmide, réprime la synthèse des sous-unités β et β' provenant du gène *rpoBC* chromosomique (Bedwell & Nomura, 1986). L'augmentation de la concentration des sous-unités β et β' libres réprime la traduction de *rpoBC* tandis qu'une fois assemblées pour former l'holoenzyme, c'est la transcription de *rpoBC* qui

est diminuée (Dykxhoorn *et al.*, 1996). Ce système n'est pas sans rappeler la régulation des opérons codant les protéines ribosomales, puisque ces dernières inhibent leur propre traduction lorsqu'elles ne sont pas assemblées (Nomura *et al.*, 1984).

La synthèse du cœur de l'ARN polymérase est soumise à des régulations complexes. Ces dernières assurent que l'ARN polymérase, composante majeure de la machinerie d'expression génique, soit exprimée quelles que soient les conditions environnementales.

2.1.1.2 Spécificité des facteurs sigmas

Chez *E. coli*, il existe 7 facteurs sigma, conférant à l'holoenzyme une spécificité pour la reconnaissance de certains promoteurs. Ces facteurs sigmas sont désignés à la fois par des lettres et par des nombres, correspondant à leur poids moléculaire en kilo Dalton. Ils permettent l'expression de différents sets de gènes (pour revue Ishihama, 2000). On retrouve ainsi σ^{70} (σ^D) responsable de la transcription des gènes de « ménage », σ^{54} (σ^N) des gènes impliqués dans la croissance en faible concentration d'azote, σ^{38} (σ^S) dans la survie, la réponse au stress et l'adaptation à la phase stationnaire, σ^{32} (σ^H) des gènes exprimés suite à un choc thermique, σ^{28} (σ^F) dans la chemotaxie et la motilité, σ^{24} (σ^E) dans le repliement des protéines dénaturées et σ^{18} (σ^{FecI}) dans le transport du citrate de fer.

Chacun d'eux permet donc la transcription d'un set de gènes spécifiques. Ils ne sont cependant pas tous exprimés au même moment et en même quantité. Hormis pour le facteur sigma végétatif (σ^{70}), la plupart des facteurs sigma sont alternatifs, et ne sont requis que pour répondre à certaines conditions physiologiques. Ainsi, les gènes codant certains facteurs σ alternatifs peuvent être inactivés sans affecter la viabilité de la cellule, à l'exception du facteur σ^{24} dont l'inactivation entraîne une chute de la viabilité chez *E. coli* (De Las Peñas *et al.*, 1997).

Lors de la phase exponentielle de croissance, la forme la plus répandue d'holoenzyme ($\alpha 2 \beta \beta' \sigma^{70}$) est responsable de la transcription des ARN stables (ARNt et ARNr) ainsi que de la plupart des gènes dits de « ménages ». Ainsi le facteur σ^{70} est présent à hauteur de 40% du niveau d'enzyme cœur dans la cellule (Engbaek *et al.*, 1976 ; Hayward & Fyfe, 1978 ; Iwakura *et al.*, 1974). Cela a également été montré pour les facteurs σ^{54} et σ^{28} malgré leurs plus faibles concentrations. Ils représentent respectivement seulement 50%

TABLE 2.1: Affinités des différents facteurs sigma pour le cœur de l'ARN polymérase chez *E. coli* (d'après Maeda *et al.*, 2000).

Sous-unité σ	$K_d(nM)$	Affinité pour le cœur	(molécules/cellule)
$\sigma^{70}(\sigma D)$	0,26	1,0	700
$\sigma^{54}(\sigma N)$	0,30	1,55	110
$\sigma^{38}(\sigma S)$	4,26	16,4	<1
$\sigma^{32}(\sigma H)$	1,24	4,75	<10
$\sigma^{28}(\sigma F)$	0,74	2,85	370
$\sigma^{24}(\sigma E)$	2,43	9,35	<10
$\sigma^{18}(\sigma FecI)$	1,73	6,65	<1

et 10% de celle de σ^{70} (Jishage *et al.*, 1996). Néanmoins, pour d'autres facteurs sigma leur niveau d'expression varie considérablement selon l'état physiologique. Le facteur σ^{38} , par exemple, voit son expression augmenter de 30% lors de l'entrée en phase stationnaire (Jishage *et al.*, 1996 ; Piper *et al.*, 2009), tout comme l'expression de σ^{32} qui est induite par un choc thermique.

2.1.1.3 Compétition entre facteurs σ

En plus de la différence de concentration intracellulaire, un autre facteur détermine la probabilité pour un facteur sigma de se lier à l'enzyme cœur : c'est sa propre affinité pour le cœur. Des expériences *in vitro* utilisant de l'enzyme cœur et des sigmas purifiés d'*E. coli* ont montré que le facteur végétatif σ^{70} a la plus forte affinité pour l'enzyme cœur et que le facteur σ^{38} est celui qui a la moins bonne affinité (Maeda *et al.*, 2000 ; Tableau 2.1). Néanmoins, pour certains facteurs σ l'affinité n'est pas le seul paramètre important pour fixer l'enzyme cœur. C'est par exemple le cas de σ^{54} qui même une fois fixé à l'enzyme cœur a besoin de facteur NtrC pour initier la transcription.

Ce concept de compétition entre les différents facteurs σ pour une quantité limitante du cœur de l'ARN polymérase a été prouvé expérimentalement, en faisant varier le niveau d'expression des différents facteurs sigma ou en générant des mutants ayant des affinités plus faibles. Par exemple, dans un mutant du facteur σ^{70} , dont l'affinité pour le cœur de l'ARN polymérase a été affaiblie, les gènes transcrits par σ^{32} , lors d'un choc thermique,

sont plus fortement exprimés que dans une souche sauvage. Il a ainsi été montré que σ^{32} ne peut gagner la compétition face à σ^{70} qu'en présence d'une quantité limitée d'enzyme cœur (Zhou & Gross, 1992). Une étude similaire, réalisée *in vivo*, a montré que le gène *uspA* reconnu par σ^{70} est plus fortement induit en l'absence du facteur σ^{38} . A l'inverse, le gène *uspB*, transcrit par ce facteur alternatif, est réprimé. En surexprimant σ^{70} , on obtient le même effet que la délétion de σ^S , alors que la surexpression de σ^{38} entraîne l'effet opposé : la répression de *uspA* et l'induction du gène *uspB* (Farewell *et al.*, 1998). Ces résultats montrent que les facteurs sigmas sont en compétition pour une quantité limitante d'ARN polymérase.

2.1.1.4 Modulation de l'affinité du facteur σ^{70}

Avec chacun leurs régulons – set de gènes cibles transcrits – les facteurs sigmas permettent à la cellule de répondre efficacement et rapidement aux changements d'environnement. Cependant, si une augmentation de la concentration de certains facteurs alternatifs a pu être observée, notamment au passage de la phase exponentielle vers la phase stationnaire Nyström (2004), elle ne suffit pas, seule, à compenser leur faible affinité pour l'ARN polymérase. Pour que les facteurs sigma alternatifs puissent concurrencer plus efficacement le facteur σ^{70} pour leur association avec l'ARN polymérase cœur, les cellules bactériennes ont développé plusieurs stratégies permettant d'en moduler l'affinité et ce, de façon réversible.

Parmi les différents effecteurs de cette modulation on peut citer (p)ppGpp, DksA, Rsd ou encore l'ARN 6S qui facilitent la redistribution de l'ARN polymérase lors d'un stress nutritionnel. En général, (p)ppGpp et DksA augmentent la transcription des gènes dépendant des facteurs alternatifs. Tandis que pour les promoteurs dépendant de σ^{70} , certains sont négativement régulés par le ppGpp (par exemple, le promoteur P1 des ARN stables) et d'autres sont positivement régulés (biosynthèse des acides aminés). Ces modulations se font par interaction directe avec l'ARN polymérase (Paul *et al.*, 2004 ; 2005 ; Gourse *et al.*, 2006 ; Potrykus & Cashel, 2008). A l'opposé, l'anti-facteur σ^{70} Rsd agit en séquestrant le facteur sigma libre et sa surexpression favorise la transcription de gènes transcrits par des facteurs alternatifs tels que σ^{24} , σ^{32} , σ^{54} et en particulier par σ^{38} (Jishage *et al.*, 2002 ; Laurie *et al.*, 2003 ; Costanzo *et al.*, 2008). Le petit ARN 6S, quant à lui, vient titrer ce facteur sigma pendant la phase stationnaire, en imitant la structure d'un promoteur σ^{70} ,

(Wassarman, 2007).

2.1.1.5 Crl : modulateur de l'activité de σ^{38}

De nombreux mécanismes mettant en évidence la compétition entre les différents facteurs sigmas ont été étudiés. Pour la plupart, ils agissent soit en diminuant l'affinité σ^{70} pour le cœur de l'ARN polymérase (par exemple Rsd), soit en donnant un avantage collectif aux facteurs sigmas alternatifs lors de différents stress notamment nutritionnel. Mais il existe aussi des mécanismes permettant de favoriser un facteur alternatif parmi les autres, c'est par exemple le cas de Crl pour le facteur σ^{38} . Cette petite protéine a tout d'abord été identifiée comme un activateur des gènes codant la formation des curli, des structures extra-cellulaires permettant l'adhérence (Arnqvist *et al.*, 1992). Plus tard, son rôle a été étendu et Crl a été présentée comme un auxiliaire de l'holoenzyme $E\sigma^{38}$ (Pratt & Silhavy, 1998). Plus récemment, il a été montré que Crl se fixe spécifiquement au facteur σ^{38} libre, favorise l'expression de certains promoteurs RpoS-dépendants à basse température et qu'il augmente probablement son affinité pour l'enzyme cœur (Bougdour *et al.*, 2004; Robbe-Saule *et al.*, 2008; England *et al.*, 2008). Plusieurs études ont montré que Crl régule tout ou partie du régulon RpoS (Lelong *et al.*, 2007; Typas *et al.*, 2007). Ainsi Crl, l'ARN 6S et l'anti- σ^{70} aident le facteur σ^{38} dans la compétition pour le cœur de l'ARN polymérase durant la phase stationnaire.

2.1.2 Ribosomes

Le ribosome est un complexe ribonucléoprotéique responsable de la traduction des ARN messagers en protéines. Les premières études ont cherché à déterminer les composants du ribosome. Ainsi, plusieurs sous-unités ont pu être isolées par ultracentrifugation et nommées en fonction de leur coefficient de sédimentation. Il est admis que le ribosome est une particule de 70S formée d'une petite sous-unité (30S) et d'une grande sous-unité (50S). De plus, deux ribosomes peuvent s'assembler pour former une particule 100S (Tissières & Watson, 1958). La petite sous-unité contient 21 protéines ribosomales (appelées S1 à S21) ainsi que de l'ARN ribosomal 16S (1542 nucléotides). La grande sous-unité est quant à elle composée de 33 protéines (L1 à L36) et de deux ARN ribosomaux, le 23S (2904 nucléotides) et le petit ARN 5S de seulement 120 nucléotides. L'ARN messager passe à

travers la petite sous-unité (30S) qui contient les sites de fixation des ARNt et permet la reconnaissance du codon de l'ARNm. La grande sous-unité contient la région catalytique, appelée centre peptidyl-transférase, qui procède à la liaison peptidique entre les acides aminés. La grande sous-unité possède également trois sites (A ou site Aminoacyl, P ou site Peptidyl et E ou site Exit) où vont se fixer les ARNt porteurs des acides aminés pendant la traduction. Le site P est occupé par un ARNt dont l'acide aminé est lié à la chaîne protéique naissante. L'ARNt portant l'acide aminé en attente d'être lié, correspondant au prochain codon, est positionné dans le site A. Enfin, le site E permet la libération de l'ARNt dé-acétylé, c'est-à-dire non-chargé de son acide aminé.

2.1.2.1 Organisation et régulation des opérons ribosomaux

Chez *E. coli*, il y a sept opérons d'ARN ribosomaux nommés *rrnA* à *rrnH*, également appelés ARN stables du fait de leur faible taux de dégradation. Leurs séquences sont presque identiques et sont donc soumises aux mêmes régulations (Condon *et al.*, 1992). Ils codent les ARN 16S, 23S, 5S et un ou plusieurs ARN de transfert (Fig. 2.4). Quatre d'entre eux sont localisés près de l'origine de répllication du chromosome *oriC*, les trois autres en sont un peu plus éloignés (Fig. 2.3). Leurs séquences cumulées ne représentent que 38kbp soit moins de 1% du génome (qui fait 4,6 Mpb). En d'autres termes, les opérons *rrn* sont en compétition avec les 99% restant du génome pour ce qui est du recrutement de l'ARN polymérase en réponse aux changement environnementaux (Jin *et al.*, 2011 ; Fig. 2.3).

Du fait de leur localisation proche de l'origine de répllication, l'expression des opérons *rrn* est affectée par le nombre de fourches de répllication, cet effet est connu sous le nom de « gene dosage effect ». En effet, chez *E. coli* le temps répllication du chromosome est de 40 min, et le temps de division cellulaire dépend du taux de croissance et peut atteindre seulement 20 min. Ainsi lorsque les cellules se divisent rapidement, on observe plusieurs fourches de répllication (Bremer & Dennis, 1996). On comprend alors que plus une cellule va posséder de fourches de répllication plus le nombre de copies des opérons *rrn* sera important. Il a été estimé qu'avec un temps de doublement de 24 min, *E. coli* présentait l'équivalent de 36 copies d'opérons *rrn*, contre 20 copies avec un temps de doublement de 40 min (Bremer & Dennis, 1996).

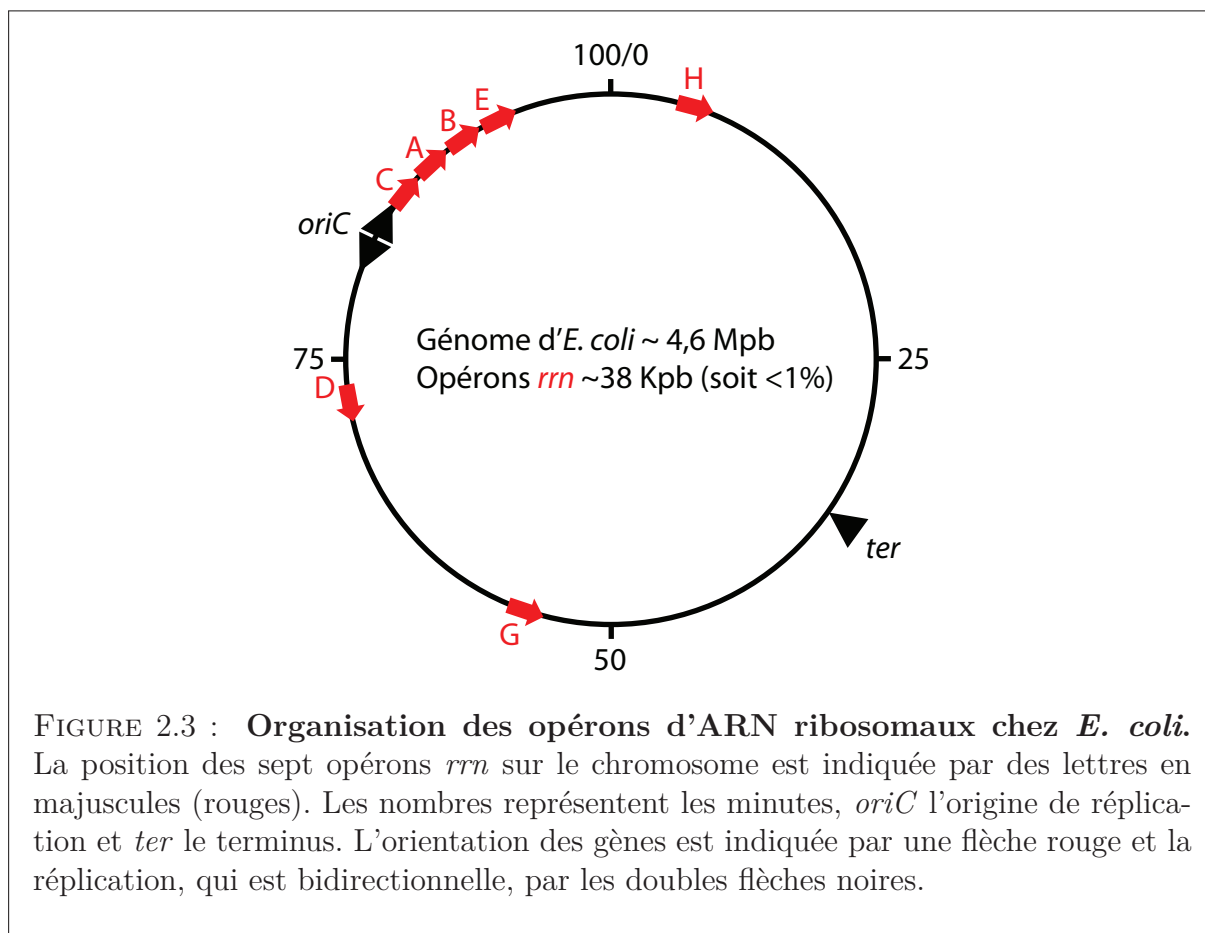


FIGURE 2.3 : Organisation des opérons d'ARN ribosomiaux chez *E. coli*.

La position des sept opérons *rrn* sur le chromosome est indiquée par des lettres en majuscules (rouges). Les nombres représentent les minutes, *oriC* l'origine de réplication et *ter* le terminus. L'orientation des gènes est indiquée par une flèche rouge et la réplication, qui est bidirectionnelle, par les doubles flèches noires.

En plus de leur localisations particulière, les opérons *rrn* ont des spécificités propres à leur structure (Fig. 2.4). Ils possèdent 2 promoteurs, P1 et P2, qui ont la même activité *in vitro* (Gilbert *et al.*, 1979 ; Glaser & Cashel, 1979) mais qui ne se comportent pas de la même façon *in vivo*.

P1 est le promoteur le plus actif, lorsque le milieu permet une croissance rapide. Il est également soumis à la réponse stringente, une régulation dépendante de la croissance (Lund & Dahlberg, 1979 ; Sarmientos & Cashel, 1983). Cette forte activité est due à la présence d'un UP élément, une séquence nucléotidique interagissant avec la sous-unité alpha de l'ARN polymérase. Cette séquence augmente l'affinité du promoteur avec l'ARN polymérase (Gourse *et al.*, 1986). Le promoteur P2, quant à lui, va plutôt interférer avec le promoteur P1, en recrutant l'ARN polymérase et en empêchant celle-ci de s'y fixer, c'est la « promoter occlusion » (Adhya & Gottesman, 1982). Une autre caractéristique, contradictoire avec sa forte expression, est la grande instabilité des complexes ouverts qu'ils forment tant que les ponts phosphodiester reliant les premiers nucléotides ne sont

pas formés (Gourse, 1988).

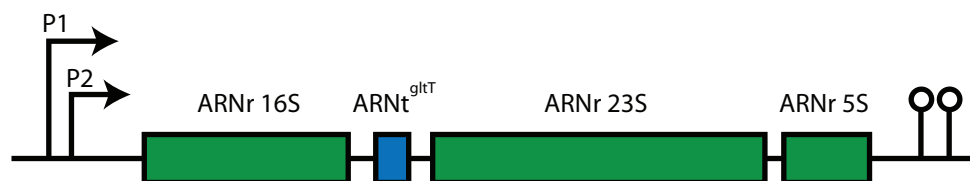


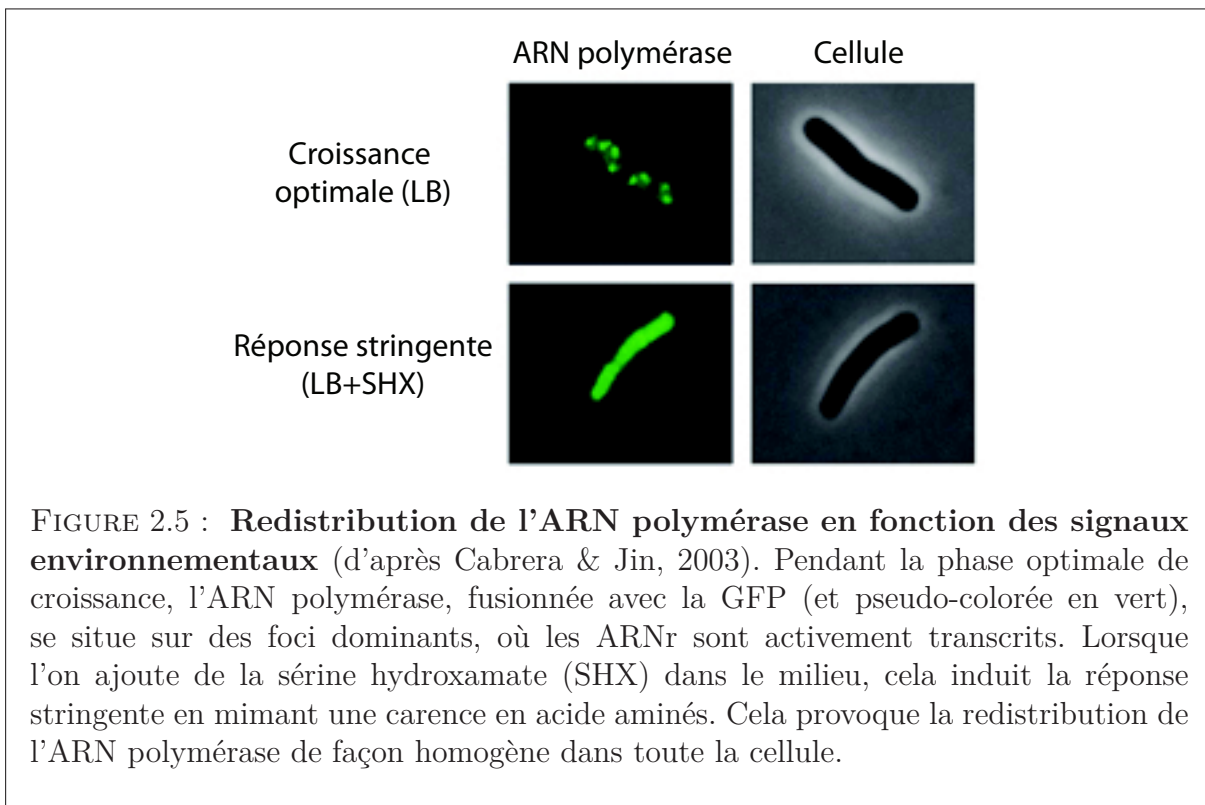
FIGURE 2.4 : **Structure générale de l'opéron *rrnB* d'*E. coli*.** Les régions qui codent les ARNr matures sont en vertes. Entre l'ARN 16S et l'ARN 23S se trouve l'ARNt *gltT*, qui est l'un des quatre ARN de transfert fixant le glutamate (représenté en bleu). Les promoteurs P1 et P2 sont schématisés par des flèches indiquant le sens de transcription. Les deux boucles noires représentent les terminateurs de la transcription de l'opéron.

Les 52 protéines ribosomales sont codées par une vingtaine d'unités transcriptionnelles localisées à 14 positions différentes sur le chromosome d'*E. coli* (Bachmann, 1990). Leur expression est majoritairement contrôlée au niveau traductionnel. L'accumulation de protéines ribosomales libre inhibe la traduction et diminue la demi-vie de leurs ARN messagers (Keener & Nomura, 1996). C'est pour cette raison que la synthèse des ribosomes dépend essentiellement du contrôle de l'expression des ARN stables. Les gènes de protéines ribosomales sont pour la plupart situés à proximité des opérons *rrn*, ce qui facilite la production et l'assemblage des ribosomes (Nomura *et al.*, 1984). Le sens de transcription des opérons *rrn* est le même que celui de la réplication de l'ADN (Ellwood & Nomura, 1982), ce qui a pour conséquence de limiter la collision entre les deux machineries (Brewer, 1988).

2.1.2.2 Réponse stringente et rôle de ppGpp

Les caractéristiques des opérons *rrn*, ainsi que leur organisation font qu'ils sont fortement exprimés et rapidement régulés en réponse au changement des conditions de croissance. Ainsi, Cashel (1996) a montré que le promoteur P1 présente une transcription quasiment inexistante lorsque les cellules n'ont plus de nutriment. De plus, les opérons *rrn* ont montré une expression différente selon la présence ou l'absence d'acides aminés dans le milieu (Stent & Benner, 1961). Lorsqu'une souche sauvage est en présence de milieu riche, contenant des acides aminés en grande quantité, la synthèse des ARN ribosomaux

est extrêmement forte. Mais dès lors qu'un seul des 20 acides aminés vient à manquer, l'expression des ARN stables s'arrête. Stent et Brenner ont appelé ce phénomène la réponse stringente. Cette chute de la transcription des opérons *rrn* entraîne d'ailleurs une redistribution de l'ARN polymérase vers d'autres promoteurs, comme les opérons de biosynthèse des acides aminés. Cette redistribution est d'ailleurs visible par microscopie, en présence d'acides aminés l'ARN polymérase forme des foci de transcription qui disparaissent lorsqu'un acide aminé vient à manquer dans le milieu (Fig. 2.5).



C'est en cherchant à relier la chute de la synthèse des ARN ribosomiaux, lors de la réponse stringente, avec la diminution de nucléoside triphosphate que Cashel et Gallant ont découvert deux nouveaux nucléotides qu'ils appelèrent : magic spot I et II. Plus tard, ils ont été identifiés comme des guanosines tetra et penta phosphate (ppGpp et pppGpp). Il a également été montré qu'en l'absence d'au moins un acide aminé, ces nucléotides sont formés seulement dans une souche possédant le gène *relA* (Cashel & Gallant, 1969).

Le gène *relA* code une pentose phosphate synthetase (PSI) qui en présence d'ARN de transfert déacétylé, c'est-à-dire non chargé d'un acide aminé, convertit du GTP en pppGpp. Celui-ci est à son tour converti rapidement en ppGpp par une pppGpp-5'-

phosphohydrolase (Haseltine *et al.*, 1972 ; Hara & Sy, 1983). Cependant, un mutant *relA* possède toujours un niveau basal de ppGpp. Ceci amena à la découverte d'une deuxième pentose phosphate synthétase (PSII) codée par le gène *spoT*. Cette enzyme possède à la fois l'activité synthétase et hydrolase du (p)ppGpp (Hernandez & Bremer, 1991). Son activité est également sensible au taux de croissance, puisque dans un mutant *relA*, le niveau de ppGpp varie avec le taux de croissance (Ryals *et al.*, 1982). Il a d'ailleurs été montré, en mesurant la synthèse et la dégradation de ppGpp au cours de la croissance et ce, dans plusieurs conditions, que l'activité de PSII est fortement instable et nécessite la production continue de protéines (Murray & Bremer, 1996).

Ces observations suggèrent que le ppGpp est impliqué dans le contrôle de la synthèse des ARN ribosomiques. En effet, le ratio entre les ARN stable et les ARN totaux diminue au fur et à mesure que le niveau de ppGpp augmente (Bremer *et al.*, 1987). Il a été montré plus récemment que le ppGpp ne répond pas uniquement à la privation d'acides aminés, mais agit également comme un régulateur global de la réponse au stress nutritionnel (Potrykus & Cashel, 2008), par exemple, lors d'un manque de sources carbonées, d'azote, de phosphate, de fer ou encore lors d'une synthèse faible d'acide gras (Srivatsan & Wang, 2008 ; Fig. 2.6).

Le ppGpp inhibe directement la synthèse des ARN stables mais cette régulation reste complexe. Il semblerait que l'effet du ppGpp sur les promoteurs *rrn* implique la protéine DksA qui interagit avec l'ARN polymérase. En effet, l'inhibition des *rrn* par ppGpp semble réduite dans un mutant *dksA* (Paul *et al.*, 2004). Néanmoins, les détails mécanistiques restent incertains car plusieurs sites d'interaction avec l'ARN polymérase ont été rapportés (Paul *et al.*, 2005 ; Haugen *et al.*, 2008 ; Potrykus & Cashel, 2008).

2.1.2.3 Assemblage et cycle de traduction

La formation des ribosomes actifs, capables de traduire un ARN messager, est un procédé complexe dont les étapes sont fortement coordonnées. On peut les résumer en 5 grandes phases : la transcription et modification des ARN ribosomiques, la traduction et modification des protéines ribosomales, la maturation et le repliement des ARNr et des protéines ribosomales, leur assemblage et la fixation de co-facteurs ainsi que leur libération. Plusieurs de ces étapes sont couplées et se déroulent de façon simultanée avec la transcription

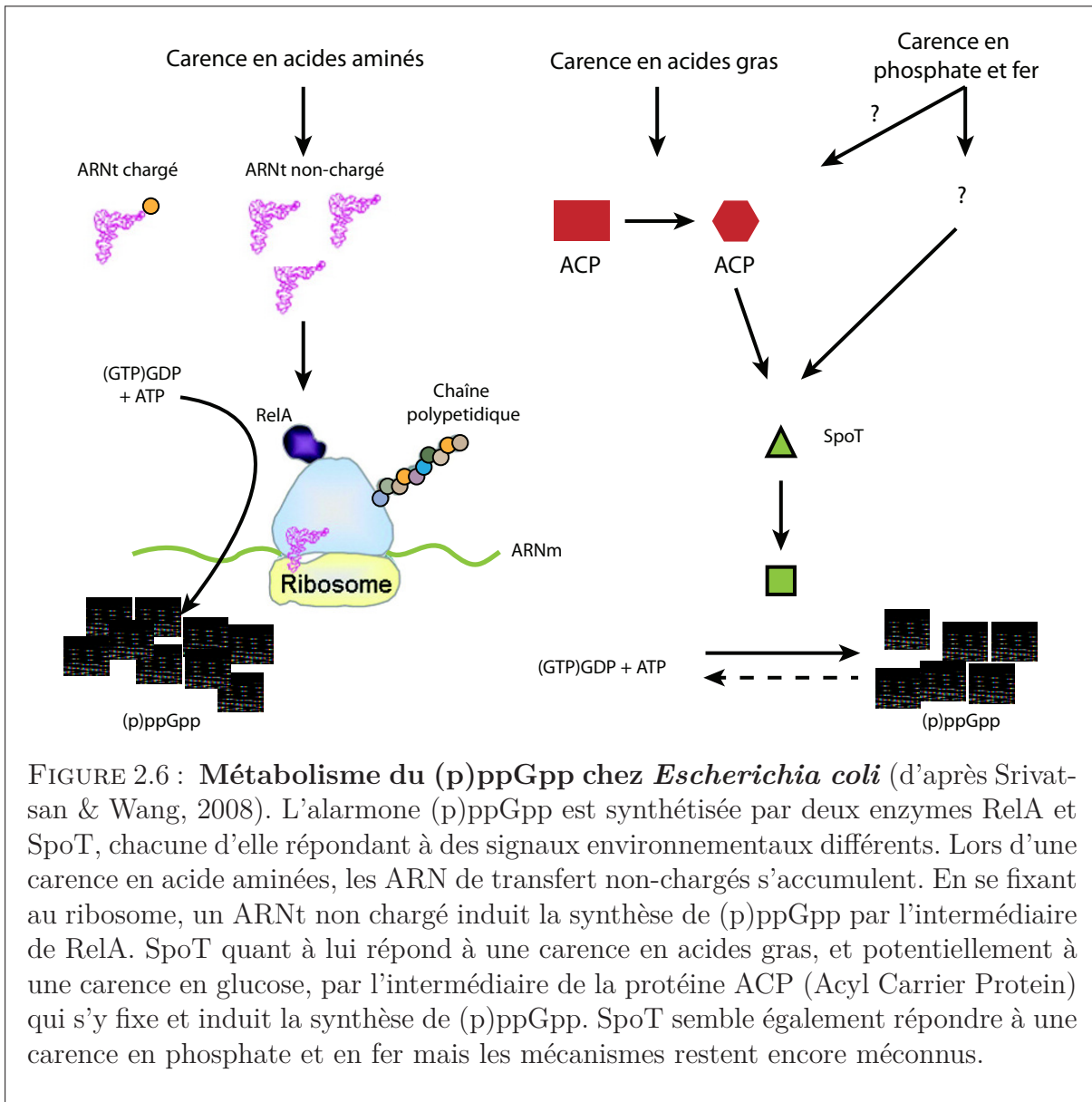


FIGURE 2.6 : Métabolisme du (p)ppGpp chez *Escherichia coli* (d'après Srivatsan & Wang, 2008). L'alarmone (p)ppGpp est synthétisée par deux enzymes RelA et SpoT, chacune d'elle répondant à des signaux environnementaux différents. Lors d'une carence en acides aminés, les ARN de transfert non-chargés s'accumulent. En se fixant au ribosome, un ARNt non chargé induit la synthèse de (p)ppGpp par l'intermédiaire de RelA. SpoT quant à lui répond à une carence en acides gras, et potentiellement à une carence en glucose, par l'intermédiaire de la protéine ACP (Acyl Carrier Protein) qui s'y fixe et induit la synthèse de (p)ppGpp. SpoT semble également répondre à une carence en phosphate et en fer mais les mécanismes restent encore méconnus.

des ARN ribosomaux. L'assemblage de tous ces éléments, pour former un ribosome 70S mature et fonctionnel, semble alterner entre des changements conformationnels des ARN et la fixation de protéines ribosomales, ces dernières permettant de stabiliser les repliements des ARN. Une fois le ribosome assemblé, il est alors capable de traduire un ARNm. Ce processus est décomposé en plusieurs étapes : l'initiation, l'élongation, la terminaison et le recyclage des sous-unités.

L'initiation de la traduction chez les bactéries fait intervenir principalement la petite sous-unité 30S du ribosome, plus particulièrement une région de l'ARN 16S, avec la séquence Shine-Dalgarno, en amont du codon initiateur de l'ARNm. Ceci permet donc

sa reconnaissance et le recrutement du premier ARNt chargé par une méthionine et fait intervenir des facteurs d'initiation appelés IF1, IF2 et IF3. S'en suit alors la phase d'élongation, au cours de laquelle le ribosome catalyse la liaison de l'acide aminé chargé sur l'ARNt, correspondant au codon lu, avec la chaîne peptidique. Une fois que l'ARNt n'a plus d'acide aminé, il est dé-acétylé, le ribosome se transloque alors vers le codon suivant et l'ARNt est libéré. Lorsque le ribosome arrive sur le codon stop de l'ARNm, celui-ci ne correspond à aucun ARNt mais à différentes protéines appelées : facteur de libération RF1, RF2 et RF3 (pour Release Factor). Ces facteurs permettent la libération de la protéine nouvellement synthétisée et la fin de la traduction. Les deux sous-unités 30S et 50S sont alors dissociées par différents facteurs tels que RRF (Ribosome Recycling Factor) et EF-G (Elongation factor G). Le facteur IF3 se fixe alors sur la sous unité 30S, empêchant ainsi la reformation du ribosome 70S, jusqu'à l'initiation du prochain cycle de traduction (Hirokawa *et al.*, 2005).

2.1.2.4 Hibernation des ribosomes en phase stationnaire

Les ribosomes peuvent représenter jusqu'à 45% de la masse totale d'une bactérie, comme *E. coli* lors de la phase exponentielle de croissance. Pendant la phase stationnaire, la biosynthèse des ribosomes est réprimée et la synthèse protéique est également fortement diminuée. Leur biosynthèse représentant un coût non négligeable pour la cellule, la répression de la traduction ne passe pas par une dégradation des ribosomes mais par leur « hibernation ». Ce stade correspond à la dimérisation des ribosomes 70S, en une particule 100S, inapte à la traduction et protégeant les ribosomes de la dégradation (Wada *et al.*, 1995). Cette dimérisation étant réversible, il est alors plus facile pour la cellule de re-mobiliser les ribosomes en cas de reprise de croissance, plutôt que de les synthétiser *de novo*.

La particule 100S est formée par la fixation sur les ribosomes de plusieurs facteurs (Fig. 2.7) : RMF (Ribosome Modulation Factor) et HPF (Hibernation Promoting Factor). RMF inactive les ribosomes en se fixant sur leur centre actif, la peptidyl-transférase (Yoshida *et al.*, 2002 ; 2004). Ce processus assure une meilleure survie de la cellule lors de la phase stationnaire. En effet, une souche mutée pour le gène *rmf* ne forme pas de particules 100S et possède une plus forte mortalité lors de la phase stationnaire comparée

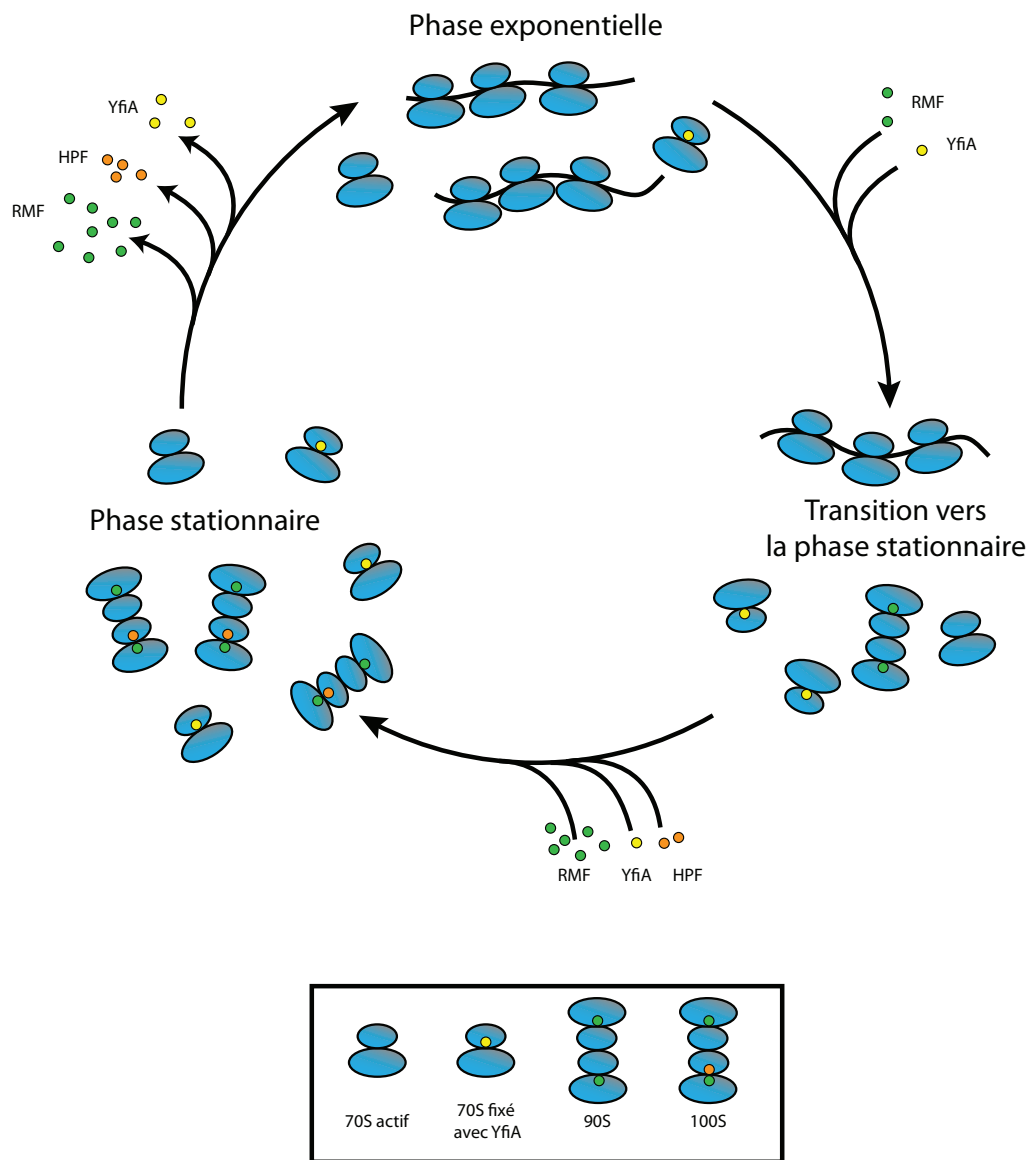


FIGURE 2.7 : **Cycle d'hibernation des ribosomes au cours de la croissance** (adaptée de Ueta *et al.*, 2005). Lors de la phase exponentielle de croissance, les ribosomes sont sous leur forme 70S et capables de traduire. La petite sous-unité 30S est schématisée par une petite ellipse et la grande sous-unité 50S par une ellipse plus large. Lors de la phase stationnaire, les ribosomes sont dimérisés sous forme de particule 100S, dans lesquelles on retrouve 2 ribosomes 70S. Les protéines RMF et HPF sont des facteurs permettant l'hibernation des ribosomes. On observe une phase intermédiaire où les ribosomes n'ont pas de protéine HPF et se présentent alors sous une taille de 90S. Que ce soit sous leur forme 90S ou 100S, les ribosomes sont alors inaptes à traduire. Lorsque l'on passe de la phase stationnaire vers une reprise de croissance, en ajoutant du milieu frais, les facteurs RMF et HPF sont alors relâchés et les ribosomes retrouvent à nouveau leur forme 70S. La protéine RMF se fixe sur la sous-unité 50S, la protéine HPF sur la petite sous unité 30S. C'est également le cas pour YfiA qui entre en compétition avec HPF et favorise le maintien des ribosomes sous leur forme 70S.

à une souche sauvage (Yamagishi *et al.*, 1993). Plus récemment les rôles de ces facteurs ont été éclaircis. Ainsi RMF, lorsqu'il est seul, permet seulement la formation de particules 90S et c'est seulement lorsque HPF vient s'y fixer que l'on passe à un dimère de ribosomes inactifs de 100S (Ueta *et al.*, 2005). Une autre protéine, YfiA, partageant 40% d'homologie avec le facteur HPF, interagit avec les ribosomes. Elle est principalement localisée sur les particules 70S, contrairement à HPF, et prévient la fixation de RMF : un mutant *yfiA* permet une fixation plus importante de RMF (Ueta *et al.*, 2005). De plus, en se fixant à la sous unité 30S, elle prévient la dissociation des ribosomes 70S et par là même, réduit le taux d'erreur lors de la traduction (Agafonov *et al.*, 1999 ; Agafonov & Spirin, 2004).

L'hibernation des ribosomes est un phénomène réversible. Ainsi, lorsque l'on place des cellules en phase stationnaire, dans du milieu frais, on observe une libération des facteurs et des ribosomes 70S réapparaissent. La libération de RMF, HPF et YfiA est extrêmement rapide, inférieure à la minute (Aiso *et al.*, 2005), et permet d'avoir des ribosomes actifs sans avoir besoin de les synthétiser. Ce mécanisme d'hibernation n'est toutefois pas complètement compris et de nombreuses études de Wada *et al.* (1995) ont été faites uniquement *in vitro*. Néanmoins, ils ont pu dresser un modèle expliquant les différents changements conformationnels des ribosomes au cours des grandes phases de croissance, présenté dans la figure 2.7.

2.1.3 Distribution de l'ARN polymérase et des ribosomes

La régulation des ARN ribosomaux repose en partie sur la compétition de l'ARN polymérase pour sa fixation entre les *rrn* et les autres gènes, selon les conditions de croissance. En outre, cette compétition indique que l'ARN polymérase est limitante dans la cellule. Chez *E. coli*, on estime la quantité d'ARN polymérase entre 2000 et 5000 molécules par génome (Ishihama, 2000 ; Grigorova *et al.*, 2006 ; Piper *et al.*, 2009).

Le taux de synthèse des ARN stables influence fortement la localisation de l'ARN polymérase entre les opérons *rrn* et les autres gènes du génome. Cette titration par les opérons *rrn* entraîne de profondes variations sur la régulation globale des gènes et sur la structure du nucléoïde en réponse aux conditions de cultures (Cabrera & Jin, 2006 ; Fig. 2.5). Ainsi, durant une croissance rapide et optimale, l'expression des opérons *rrn* occupe la majorité des ARN polymérases, ce qui diminue fortement leur disponibilité pour

transcrire les autres gènes. A l'inverse, lors de la réponse stringente il y a beaucoup plus d'ARN polymérase libres et disponibles pour transcrire les gènes autres que ceux codant les ARN stables.

Plusieurs modèles expliquant l'effet du ppGpp sur l'ARN polymérase ont été proposés. Selon le premier modèle, l'ARN polymérase est partitionnée en deux formes : une forme interagissant avec ppGpp qui est incapable d'initier la transcription des promoteurs *rrn* mais celles des autres promoteurs, et une forme sans ppGpp capable d'initier la transcription de l'ensemble des promoteurs, opérons *rrn* compris (Baracchini & Bremer, 1988). Le deuxième modèle considère que l'ARN polymérase, en présence de ppGpp, va faire des pauses lors de l'élongation. Ceci entraîne une diminution de sa disponibilité pour l'expression globale des gènes (Jensen & Pedersen, 1990).

2.2 Régulation spécifique de l'expression génique

Au cours de l'adaptation de l'expression génique à des changements environnementaux, un deuxième niveau de régulation s'ajoute à celui exercé par la machinerie globale d'expression génique. Il fait intervenir des protéines qui régulent spécifiquement l'expression de sets de gènes en fonction des conditions environnementales. Nous nous focalisons ici sur les principaux mécanismes de régulations transcriptionnelles auxquels je me suis intéressé pendant ma thèse. Nous ne traiterons donc pas des autres mécanismes connus comme la régulation de la traduction et la dégradation.

2.2.1 Régulation par des facteurs de transcription

La régulation transcriptionnelle provient de l'interaction entre un facteur en *trans* (molécule ou protéine qui fixe l'ADN) et un élément en *cis* (une région d'ADN). Un exemple de région régulatrice en *cis* est l'opérateur *lacO* de l'opéron lactose. Le répresseur LacI vient se fixer sur cette séquence, et vient réprimer ainsi la transcription de l'opéron *lacZYA*. L'opérateur quant à lui ne code aucun ARN ou protéine. Chez les bactéries, ces régions peuvent aller jusqu'à 400 paires de bases en amont du site d'initiation de la transcription (Collado-Vides *et al.*, 1991).

En se basant sur deux de leurs domaines de régulation, il est possible de classer les facteurs de transcription en plusieurs familles (Jacob, 1970). Le premier de ces domaines fonctionnels correspond à la détection d'un signal par la liaison d'un ligand ou par une interaction protéine-protéine. Dans la plupart des cas, le ligand est un métabolite ou un signal physicochimique provenant du cytoplasme ou du milieu environnemental (Ptashne & Gaan, 2002 ; Martínez-Antonio & Collado-Vides, 2003). L'autre domaine interagit directement avec la séquence d'ADN, le plus souvent c'est un domaine hélice-tour-hélice (Madan Babu, 2003). La plupart des facteurs de transcription possèdent ces deux domaines en une seule protéine (Ulrich *et al.*, 2005). Il existe cependant un cas particulier : celui des systèmes à deux composants. Ils sont composés d'une protéine détectrice, habituellement trans-membranaire, qui s'auto-phosphoryle en présence d'un signal dans le milieu. Elle va également phosphoryler son partenaire cytoplasmique qui possède le domaine de régulation transcriptionnelle (Mascher *et al.*, 2006).

En général, les facteurs de transcription régulent positivement un gène, on parle alors d'activation, ou négativement, on parle alors de répression. L'exemple le plus connu est celui de l'opéron *lac*, permettant le transport et le catabolisme du lactose. Le répresseur LacI (Gilbert & Müller-Hill, 1966), lorsqu'il n'y a pas de lactose dans le milieu, va se fixer sur les opérateurs *lacO* de la région promotrice de l'opéron *lacZYA* et empêcher sa transcription. En présence de lactose, LacI se fixe aux molécules de lactose, il possède une plus grande affinité avec le lactose qu'avec les opérateurs *lacO*. Ceci permet la transcription de l'opéron et l'utilisation du lactose par la cellule comme source de carbone.

Mais certains de ces facteurs de transcription peuvent avoir les deux effets : activateur et répresseur. C'est en général le cas pour les régulateurs des opérons permettant le catabolisme des sucres. Prenons l'exemple de l'opéron arabinose. En l'absence d'arabinose dans le milieu, le régulateur AraC se fixe sur l'opérateur *araO* et réprime l'opéron *araBAD* (qui code les transporteurs et enzyme de dégradation de l'arabinose). Mais dès lors qu'il y a de l'arabinose dans le milieu, AraC s'y complexe, change de conformation et active la transcription de *araBAD* en se fixant sur *araI*, un autre opérateur.

Les régulateurs transcriptionnels régulent l'expression de gènes dont certains codent aussi des régulateurs. Ces interactions forment ce que l'on appelle un réseau de régu-

lation génique. Chez les génomes bactériens, le nombre de régulateurs transcriptionnels augmente avec le nombre de gènes présents sur le génome, et forment des réseaux de régulation complexes (Cases *et al.*, 2003 ; Molina & Van Nimwegen, 2008). Ces réseaux confèrent aux bactéries leur formidable capacité d'adaptation et de survie face à leur environnement. Il est d'ailleurs possible de représenter ces réseaux de régulation sous forme de graphe, comme celui d'*Escherichia coli* (Fig. 2.8).

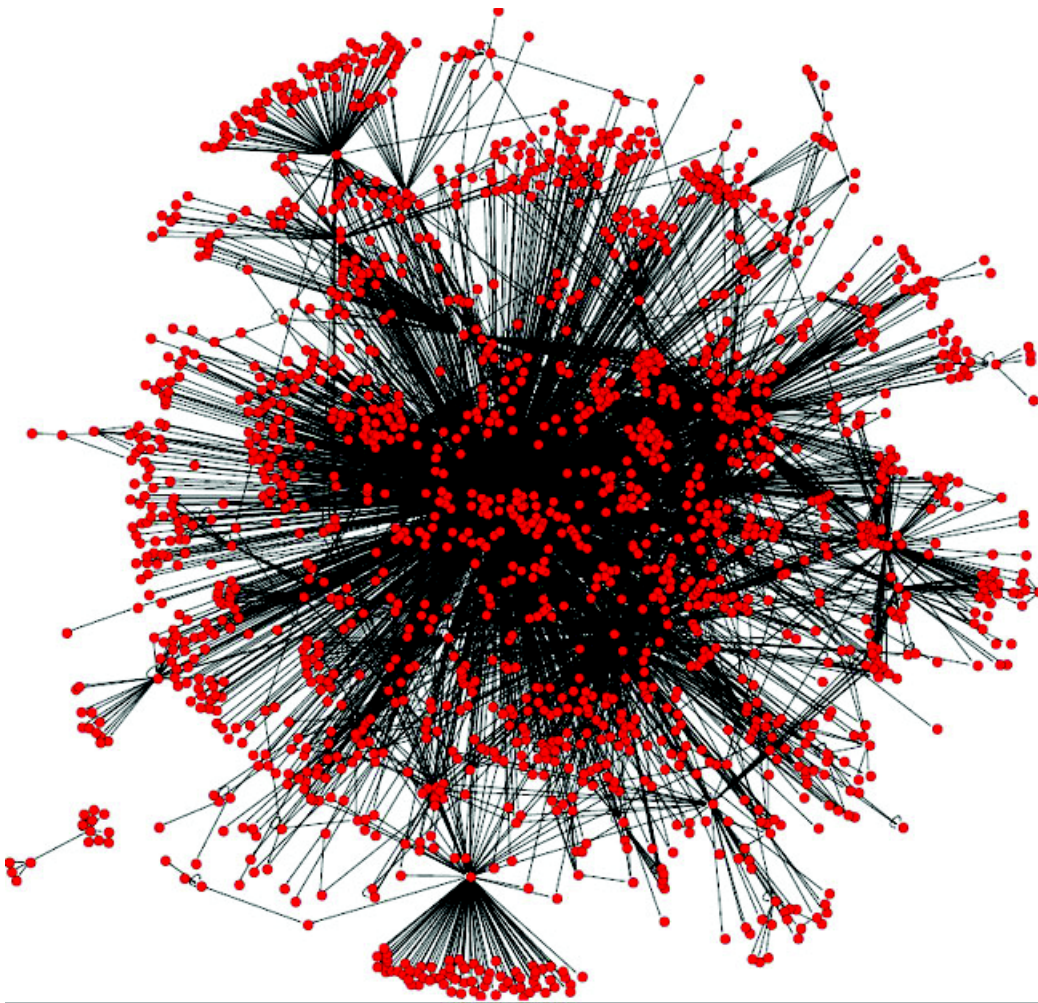


FIGURE 2.8 : Complexité du réseau de régulation transcriptionnelle d'*Escherichia coli* (source : Freyre-Gonzalez & Trevino-Quintanilla, 2010). Sur ce graphe, les gènes sont représentés par les nœuds (rouges) et les arêtes représentent les régulations qui les relient. Ce réseau a pu être construit en utilisant les données de la base RegulonDB (Salgado *et al.*, 2006).

2.2.2 Régulation par les protéines associées au nucléoïde (NAPS)

Ces dernières années, un autre facteur important dans l'expression des gènes est étudié : le rôle des protéines associées au nucléoïde. En effet, le chromosome d'*E. coli* est condensé de l'ordre de 1000 fois pour former le nucléoïde (Pettijohn, 1996). On y retrouve de nombreuses petites protéines capables de fixer l'ADN et qui forment des complexes instables avec ce dernier, pouvant recouvrir jusqu'à la moitié du chromosome (Ishihama, 1999). Il a été montré que ces protéines ont un rôle dans l'organisation et la compaction du chromosome même si les mécanismes exacts ne sont pas encore tous connus à ce jour (Dame, 2005). En plus de ce rôle structurel, plusieurs observations montrent que ces protéines régulent également la transcription.

Les protéines liant l'ADN présentes dans le nucléoïde sont nombreuses et les concentrations de certaines ont pu être mesurées, notamment pour les 12 plus abondantes lors d'une croissance rapide ou de la phase stationnaire (Azam *et al.*, 1999). Ainsi les nucléoprotéines majeures lors de la phase exponentielle sont : Fis (factor for inversion stimulation), HU, IHF (« integration host factor »), StpA (« suppressor of td mutant phenotype A ») et H-NS (« histone-like nucleoid-structuring protein »). Lors de la phase stationnaire, la composition du nucléoïde change et l'on retrouve plus particulièrement Dps (« DNA-binding protein from starved cells »), IHF et HU. Cette modification de la composition en NAPS (« nucleoid associated proteins ») laisse penser qu'elle est responsable des changements de structure et d'activité du nucléoïde. Il a été montré que les nucléoprotéines contrôlent l'expression d'une très grande variété de gènes essentiels pour la survie cellulaire, comme des gènes impliqués dans le métabolisme ou la synthèse protéique. Par exemple la protéine Fis, qui est très abondante au début de la phase exponentielle de croissance, régule positivement l'expression d'un bon nombre de gènes fortement transcrits pendant cette phase rapide de croissance, les ARN stables entre autres (Finkel & Johnson, 1993).

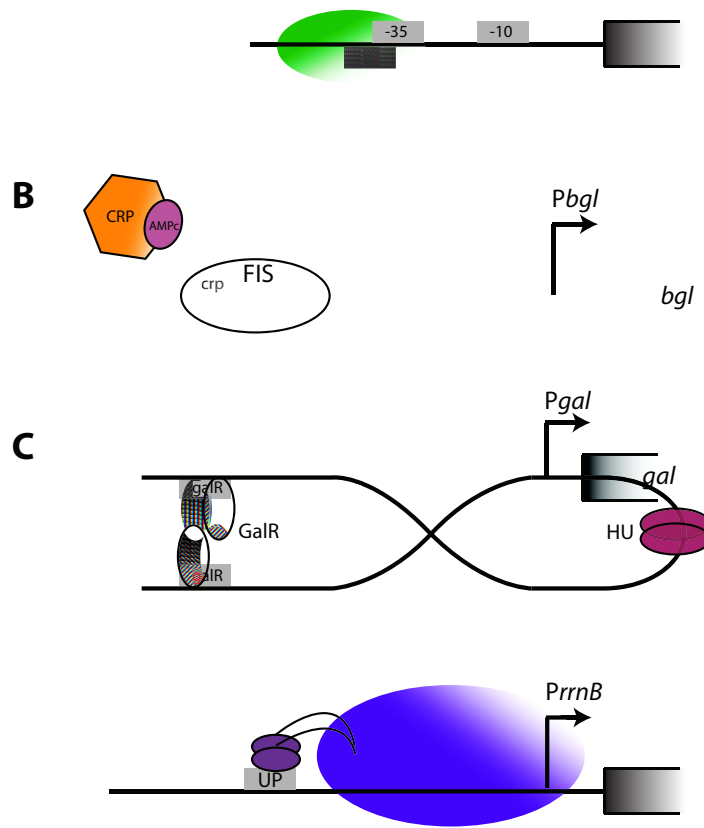
2.2.2.1 Modes de répression par les NAPS

Plusieurs mécanismes de répression des gènes par les NAPS ont été découverts, le plus répandu est l'occlusion du promoteur, lorsque la nucléoprotéine se fixe sur le promoteur du gène et empêche l'ARN polymérase de s'y fixer. C'est le cas de Fis qui réprime la plupart des promoteurs de cette façon et notamment *gyrA* (Finkel & Johnson, 1993 ; Schneider *et al.*, 1999 ; Fig. 2.9 A). Les promoteurs réprimés par Fis possèdent généralement plusieurs sites de liaison en amont et en aval du site d'initiation de la transcription, permettant ainsi une meilleure séquestration du promoteur. Il semble que la protéine H-NS agisse de la même manière sur les promoteurs qu'elle régule (Alper & Stephanopoulos, 2007).

Ces deux protéines peuvent également réprimer des promoteurs en se fixant sur ceux-ci mais sans empêcher la fixation l'ARN polymérase. C'est le cas d'H-NS sur le promoteur *rrnB* P1, l'ARN polymérase peut alors initier la transcription mais le complexe ouvert est incapable de procéder à l'élongation (Schröder & Wagner, 2000). Le gène *gyrB* est également réprimé par un mécanisme similaire, la protéine Fis se fixe alors en amont de l'ARN polymérase permettant à celle-ci de se fixer mais l'empêchant de transcrire le gène (Schneider *et al.*, 1999).

Les nucléoprotéines peuvent également réprimer l'expression de certains gènes en rentrant en compétition avec un activateur de la transcription. C'est le cas du gène *bgl*. Fis se fixe 52 paires de base en amont du site d'initiation de la transcription, empêche la fixation du complexe activateur CRP-AMPc et réprime ainsi son expression (Fig. 2.9 B). Il a d'ailleurs été montré que cette répression peut être supprimée par une mutation du promoteur augmentant l'affinité du complexe CRP-AMPc (Caramel & Schnetz, 2000).

D'autres mécanismes ont pu être mis en évidence, comme la répression indirecte d'un gène en facilitant le recrutement d'un répresseur. La nucléoprotéine HU permet au répresseur GalR de se fixer plus facilement, en modifiant la structure locale de l'opéron *gal* (Fig. 2.9 C). Ainsi en changeant seulement l'architecture de l'ADN, l'expression de *gal* est réprimée par HU (Lewis *et al.*, 1999). On peut également citer le cas du gène *proU* dont l'expression est réprimée par H-NS, en modifiant la topologie de son promoteur sans que celle-ci n'interagisse directement avec l'ARN polymérase (Jordi *et al.*, 1997).



2.2.2.2 Modes d'activation par les NAPS

La plupart des NAPS sont également capables d'activer directement ou indirectement la transcription de certains gènes. La protéine Fis peut agir comme un activateur de la transcription en interagissant directement et spécifiquement avec l'ARN polymérase. Cette interaction a été notamment bien caractérisée pour les promoteurs *rrnB* et *proP* (Bokal *et al.* (1995) ; Xu & Johnson (1995) ; Fig. 2.9 D). Plusieurs sites de fixation de Fis sont présents en amont du UP élément, un domaine riche en AT permettant la fixation du domaine α -CTD de l'ARN polymérase. Ainsi l'activation par Fis se fait en facilitant le recrutement de l'ARN polymérase sur ces promoteurs.

D'autres nucléoprotéines sont également capables d'activer la transcription sans autre facteur. Il a été montré que la courbure de l'ADN provoquée par la fixation de IHF est nécessaire pour l'activation de certains gènes, en particulier pour le gène *malT*. Des mutants de la protéine IHF permettent une courbure optimale de la région promotrice de *malT* procurant alors une activation maximale de celui-ci (Engelhorn & Geiselmann, 1998). Dans le cas du promoteur PL du phage lambda, IHF en se fixant à l'ADN courbe la région promotrice et permet à l'ARN polymérase d'interagir avec un UP élément jusqu'alors inaccessible car trop éloigné (Giladi *et al.*, 1998 ; Fig. 2.9 E). Il permet également l'activation des gènes σ^{54} -dépendants en courbant leur région promotrice pour permettre à l'ARN polymérase d'entrer en contact avec le facteur NtrC nécessaire à l'activation de cette classe de promoteur (Goosen *et al.*, 1995).

Le rôle des nucléoprotéines dans le contrôle de l'expression des gènes est de mieux en mieux compris. Malgré la bonne compréhension de leur fonctionnement, de nombreux mécanismes complexes de régulation restent à élucider, du fait qu'elles n'agissent pas toujours seules mais la plupart du temps de concert avec d'autres NAPS. Elles peuvent ainsi moduler la transcription du génome face à des changements environnementaux en réprimant ou en activant des régions entières du chromosome ou encore, en facilitant l'activité de promoteurs plus spécifiques.

Objectifs

Ma thèse a porté sur différents aspects de la régulation génique et de la croissance chez *Escherichia coli*. Mon objectif a consisté d'une part à étudier la dynamique d'expression et la régulation du gène *crl* et d'autre part, à contrôler la croissance de la bactérie en modulant l'expression de la machinerie globale d'expression des gènes.

La première partie de cette thèse est dédiée à l'étude de l'expression d'un régulateur global spécifique dans des cultures en batch d'*E. coli*, le gène *crl*. Il module l'activité du facteur sigma RpoS et partage avec lui un régulon représentant environ 10% du génome de cette bactérie. Lorsque j'ai débuté ma thèse, la régulation de son expression était très peu connue. C'est pourquoi nous avons caractérisé son profil d'expression au cours de la croissance à l'aide de fusions rapportrices. Puis, pour identifier des régulateurs responsables de ce profil, nous avons criblé une centaine de mutants de gènes impliqués dans la perception de l'environnement : gènes du quorum sensing et systèmes à deux composants, ainsi que d'autres régulateurs globaux.

Alors que je menais ces travaux, nous avons constaté au laboratoire que l'expression de la plupart des régulateurs globaux d'*E. coli* présentent un profil d'expression similaire à *crl* lors de la reprise de croissance. Les régulations spécifiques ne pouvant entièrement expliquer le profil de *crl*, nous nous sommes intéressés au rôle de la machinerie globale d'expression des gènes et son impact sur la croissance. En effet, plusieurs études ont montré que des promoteurs non régulés présentaient une expression différente selon le taux de croissance. De plus, le nombre d'ARN polymérase et de ribosomes disponibles, autrement dit « la machinerie globale d'expression », varie aussi selon le taux de croissance. Ceci suggère que la machinerie globale pourrait être à l'origine des variations d'expression

génique, lorsque le taux de croissance change.

La deuxième partie de cette thèse a consisté à vérifier cette hypothèse. Dans ce but, nous avons construit un système nous permettant de contrôler la croissance d'*E. coli* en modulant la disponibilité de sa machinerie globale. Dans un premier temps, nous avons tenté de moduler l'activité traductionnelle de la cellule, en influençant la dimérisation et donc l'inactivation des ribosomes. Cette approche n'ayant pas donné les résultats es-comptés, nous avons dans un deuxième temps modifié l'expression de l'ARN polymérase. Parmi les sous-unités qui la composent, $\beta\beta'$ sont les sous-unités les moins exprimées et donc limitantes pour former l'enzyme coeur. Nous avons donc décidé de modifier leur expression. L'opéron *rpoBC*, codant pour ces sous-unités, a donc été placé sous le contrôle d'un promoteur inductible par l'IPTG, nous permettant ainsi de contrôler le niveau de transcription et de moduler le taux de croissance de la cellule. Par la suite, nous avons caractérisé les effets de différents niveaux d'ARN polymérase sur la physiologie cellulaire, tant à l'échelle de la population qu'à celle de la cellule unique. En collaboration avec l'équipe de F. Taddei et A. Lindner, nous avons dans ce dernier cas suivi la croissance des cellules à l'aide d'un dispositif de microfluidique. Dans un dernier temps, nous avons recherché les conséquences du contrôle du niveau de l'ARN polymérase sur l'expression de l'ensemble des gènes, par séquençage d'ARN messagers.

Résultats

Étude de l'expression de régulateurs globaux : cas du gène *crl*

3.1 Mesure dynamique de l'expression des gènes

Au laboratoire, nous étudions les réseaux de régulation, c'est-à-dire comment les gènes se régulent entre eux pour produire des fonctions complexes et ainsi permettre aux cellules de s'adapter rapidement et efficacement aux changements de leur environnement.

Nous nous intéressons donc à la dynamique d'expression des gènes au cours de la croissance. L'approche que nous utilisons consiste à mesurer l'expression de gènes d'intérêt à l'aide de gènes rapporteurs tels que la luciférase ou la GFP (Green Fluorescent Protein). Cela consiste à fusionner la région promotrice et dans certains cas - une partie de - la région codante d'un gène d'intérêt à un gène rapporteur. L'expression du gène rapporteur produit un signal visible et facilement mesurable (luminescence ou fluorescence) qui reflète l'expression du gène d'intérêt. Au cours de ma thèse, j'ai essentiellement utilisé la luciférase comme système rapporteur. Nous utilisons l'opéron *luxCDABE* car il permet à la fois la production de l'enzyme (par *luxAB*) et de son substrat (par *luxCDE*). Ceci rend le système rapporteur indépendant de l'ajout de substrat, nécessaire à la réaction de bioluminescence, dans le milieu, tel que le décanal.

Une méthode courante consiste à placer le système rapporteur sur un plasmide à faible nombre de copies (voir Fig. 3.1 pour un exemple). Ainsi le signal émis est amplifié,

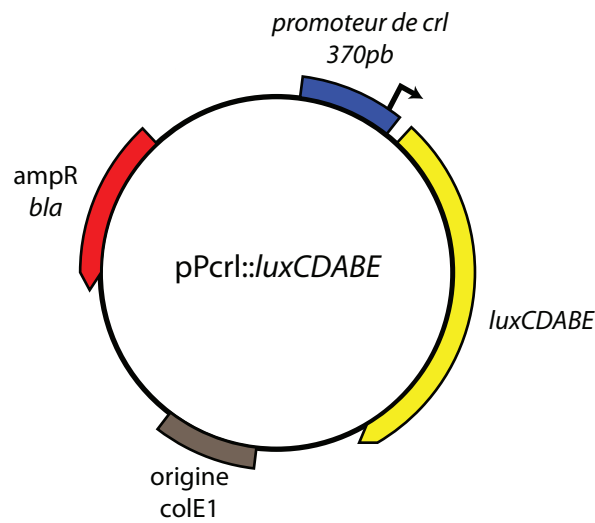
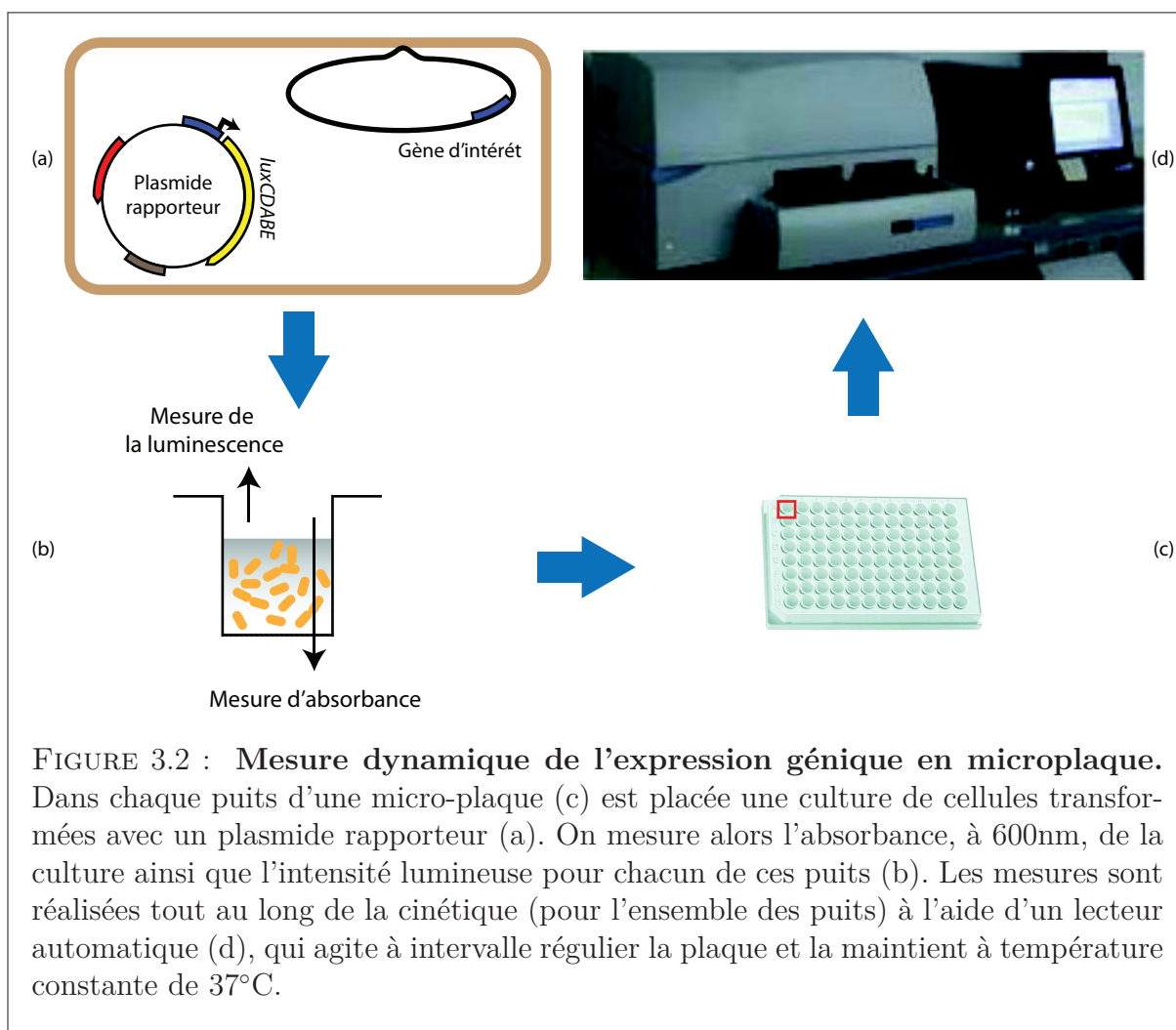


FIGURE 3.1 : **Plasmide rapporteur contenant le promoteur du gène *crl* en amont de l'opéron luciférase.** La région promotrice du gène *crl* (en bleu) a été clonée en amont de l'opéron *luxCDABE*, codant pour la luciférase (en jaune). L'origine de réplication ColE1 (en noire) régule le nombre de copie du plasmide. Le gène *bla* confère la résistance à l'ampiciline (en rouge) et sert de moyen de sélection pour les cellules transformées.

tout en limitant le coût pour la cellule. En effet, la présence d'un plasmide engendre un coût pour sa croissance, d'autant plus si le nombre de copies est important. Il faut par exemple exprimer le gène de résistance permettant de maintenir le plasmide. De plus le promoteur étudié n'est pas dans son contexte génétique local et certaines régulations peuvent être absentes, comme la topologie de la région chromosomique. Cependant, cette méthode a l'avantage d'être facile à utiliser et permet de mesurer l'expression d'un gène dans plusieurs contextes génétiques, il suffit de transformer le plasmide dans différentes souches. Il est également possible de recombinaison le système rapporteur directement sur le chromosome et ainsi s'affranchir de certains biais, comme la variation du nombre de copies de plasmides, mais dès lors que le nombre de souches à tester devient important les recombinaisons qui en découlent deviennent trop nombreuses et laborieuses.

Pour mesurer l'expression de nos systèmes rapporteurs, au cours de la croissance en milieu liquide, nous utilisons un lecteur de micro-plaque. Ce dernier peut lire l'absorbance, la fluorescence ainsi que la luminescence dans chacun des 96 puits d'une micro-plaque, toutes les 2-3 min, tout en permettant aux cultures de croître. En effet, il est possible

de thermostatier l'appareil à 37°C et d'agiter régulièrement la micro-plaque afin d'assurer une oxygénation convenable et d'éviter aux cellules de décanter.



3.2 Analyse des cinétiques d'expression

La combinaison de ces deux outils, fusion rapportrice et lecteur de microplaque automatique, nous permet de tester un grand nombre de conditions en parallèle. Une expérience typique correspond au suivi de 96 cultures en parallèle pendant plusieurs heures, avec éventuellement plusieurs fusions. Ceci produit une énorme quantité de données, jusqu'à 100000 mesures d'absorbance, de luminescence et de fluorescence par expérience. Cette grande quantité de données nécessite d'être traitée et analysée afin de pouvoir déterminer l'expression des gènes d'intérêts mesurée, comme la concentration du rapporteur par

cellule ou encore l'activité du promoteur. L'équipe a mis au point une méthode mathématique permettant de calculer de telles informations à partir des données brutes (de Jong *et al.*, 2010), je présenterai ici quelques détails de cette méthode dont j'ai fait usage au cours de ma thèse.

J'ai essentiellement utilisé la luciférase comme système rapporteur ; je vais donc décrire les différentes étapes permettant le traitement de ce type d'informations. Pour chaque expérience, nous obtenons des données brutes d'absorbance et de luminescence pour chacun des puits. Dans un premier temps, l'absorbance mesurée pour un puits contenant des cellules a besoin d'être corrigée, il faut pour cela retirer le niveau basal d'absorbance obtenu dans un puits ne contenant que du milieu de culture. L'équation (3.1) permet d'apporter cette correction et l'on obtient ainsi l'absorbance corrigée $A(t)$ en retirant l'absorbance mesurée pour le puits contrôle, $A_b(t)$, à celle mesurée pour le puits d'intérêt, $A_u(t)$.

$$A(t) = A_u(t) - A_b(t) \quad (3.1)$$

Il faut également corriger la luminescence en retirant l'intensité lumineuse émise par un puits de référence. Pour obtenir cette mesure nous utilisons un puits contrôle contenant des cellules transformées avec un plasmide rapporteur dépourvu de promoteur. Nous pourrions aussi bien utiliser de simple cellules non transformées puisque *Escherichia coli* n'émet pas naturellement de lumière. Cependant la présence d'un plasmide dans une cellule engendre toujours un coût pour cette dernière, c'est pourquoi il est plus rigoureux d'utiliser un plasmide contrôle. L'équation (3.2) décrit la méthode de correction apportée à l'intensité mesurée pour le puits d'intérêt, $I_u(t)$, auquel l'on soustrait l'intensité basale $I_b(t)$ afin d'obtenir une intensité lumineuse corrigée. Dans le cas où l'on utilise la GFP comme système rapporteur, on ne peut soustraire directement la fluorescence du puits contrôle. En effet, les cellules produisent une autofluorescence qui varie selon la taille de la population (donc de son absorbance). C'est pourquoi il faut tenir compte de la différence d'absorbance entre le puits mesuré et le puits contrôle (de Jong *et al.*, 2010).

$$I(t) = I_u(t) - I_b(t) \quad (3.2)$$

Une fois les données corrigées pour le bruit de fond, il est alors possible de calculer la concentration moyenne de luciférase dans une bactérie. En effet, l'intensité lumineuse

émise est directement proportionnelle au nombre de molécules de luciférase par cellule. Il existe cependant un biais à considérer : pour émettre un photon, la luciférase a besoin d'énergie. Dès lors que l'énergie devient limitante dans la cellule, l'activité enzymatique de la luciférase s'en voit affectée et l'intensité mesurée ne reflète plus la concentration réelle d'enzyme. Cet artefact est toutefois facile à identifier, puisque l'intensité lumineuse chute très rapidement, en quelques minutes. Un changement aussi rapide ne pourrait être expliqué par une variation soudaine de la concentration en luciférase et apparaît en général lors de la phase stationnaire prolongée, c'est-à-dire en condition de stress nutritionnel. Hormis ce cas de figure, on peut calculer la concentration de luciférase par cellule, $[Luciférase](t)$, avec l'équation (3.3), en faisant le ratio entre l'intensité corrigée d'un puits, $I(t)$, et l'absorbance corrigée de ce même puits, $A(t)$.

$$[Luciférase](t) = I(t)/A(t) \quad (3.3)$$

Un exemple de traitement de données est présenté dans la figure 3.3. Ces données concernent la fusion rapportrice du gène *crl*, codant un régulateur global dont j'ai étudié l'expression au cours de ma thèse. Les résultats de cette étude seront détaillés dans la section 3.4.

Au vu de la grande quantité de données produite par notre approche, il était nécessaire d'automatiser ce type de traitement. Pour répondre à ce besoin, nous avons conçu un programme nommé WellReader, développé en MATLAB, intégrant les fonctionnalités énoncées précédemment, telles que la soustraction du bruit de fond et le calcul de la concentration du rapporteur ; ainsi que la détection d' »outliers » (points aberrants) et le lissage des données par des fonctions « spline ». En tant qu'utilisateur quotidien de ce programme, j'ai plus particulièrement participé au développement de l'interface graphique et notamment au module permettant de superposer les données de différents puits, afin de pouvoir facilement comparer plusieurs conditions. Ce programme a fait l'objet d'une note dans le journal, *BioInformatics*, intitulée « WellReader : a MATLAB program for the analysis of fluorescence and luminescence reporter gene data ».

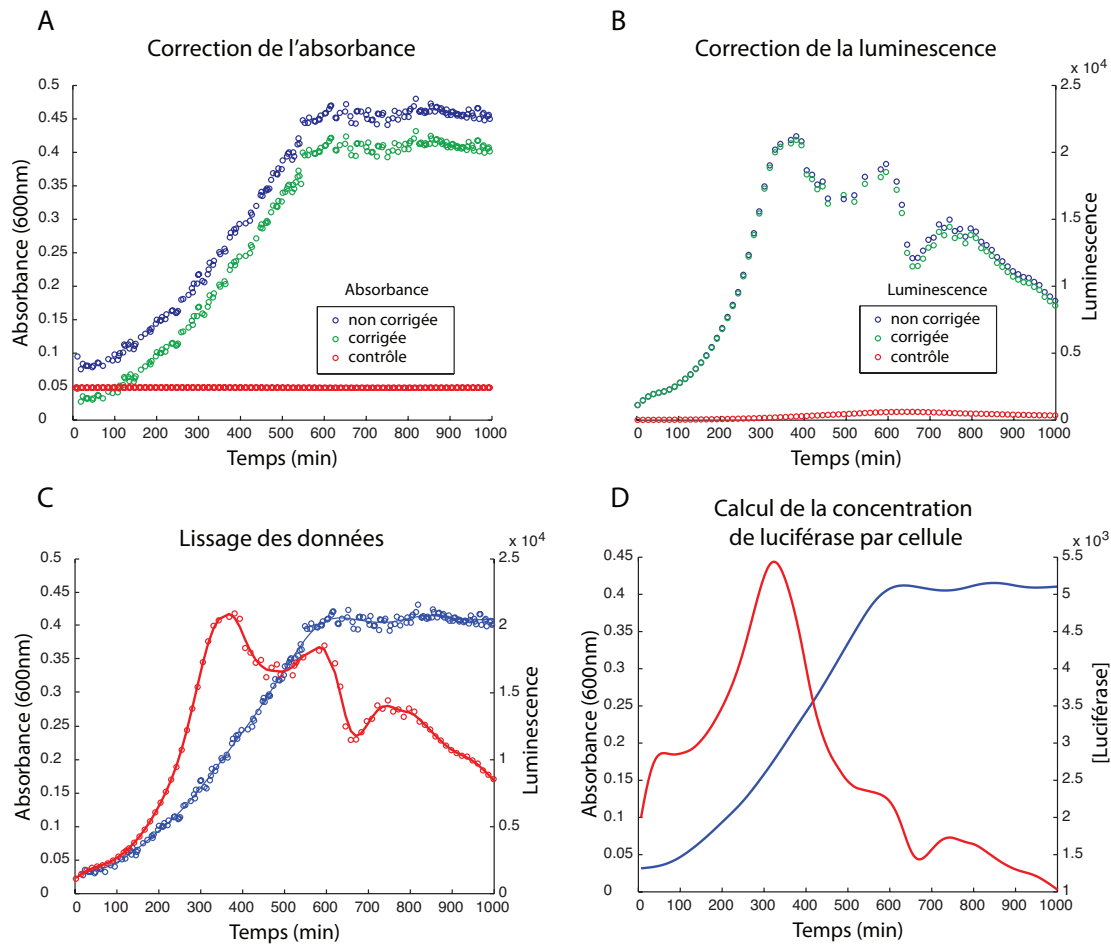


FIGURE 3.3 : Exemple de traitement de données obtenues avec la fusion *crl : lux*. (A) Les données d'absorbance brutes provenant d'une culture transformée avec la fusion *crl* (cercles bleus) sont corrigées en retirant l'absorbance d'un puits contrôle, contenant uniquement du milieu de culture (cercles rouges). On obtient alors l'absorbance corrigée en vert. Chaque cercle représente une mesure d'absorbance. (B) Les données de luminescence brutes (cercles bleus), issues de la même culture que dans la figure A, sont corrigées en retirant la luminescence d'un puits contrôle contenant le système rapporteur dépourvu de promoteur (cercles rouges). Les données corrigées sont tracées en vert. Chaque mesure de luminescence est représentée par un cercle. (C) Les données corrigées (traits pleins), en bleu pour l'absorbance, en rouge pour la luminescence, sont lissées à l'aide d'un spline polynomial. (D) A partir des données d'absorbance, de la luminescence corrigées et lissées, on peut calculer la concentration du rapporteur en utilisant l'équation 3.3. On obtient ainsi la concentration de luciférase par cellule (rouge). L'absorbance est en bleu. La chute de luminescence observée vers 700 min (en B, C et D) est un exemple d'artefact propre à l'utilisation de la luciférase. Son activité, et non sa concentration, diminue momentanément suite à la chute du pouvoir réducteur intracellulaire lors de l'entrée en phase stationnaire (Koga *et al.*, 2005). Ce phénomène est connu sous le nom d'ADLA (Abrupt Decline of Luciferase Activity).

Publication n°1 :

**WellReader : a MATLAB program for the analysis of
fluorescence and luminescence reporter gene data**

WellReader: a MATLAB program for the analysis of fluorescence and luminescence reporter gene data

Frédéric Boyer¹, Bruno Besson¹, Guillaume Baptist^{1,2}, Jérôme Izard^{1,2}, Corinne Pinel^{1,2}, Delphine Ropers¹, Johannes Geiselmann^{1,2} and Hidde de Jong^{1,*}

¹INRIA Grenoble-Rhône-Alpes, 655 avenue de l'Europe, 38334 Saint Ismier Cedex and ²Laboratoire Adaptation et Pathogénie des Microorganismes (CNRS UMR 5163), Université Joseph Fourier, Bâtiment Jean Roget, Domaine de la Merci, 38700 La Tronche, France

Associate Editor: Trey Ideker

ABSTRACT

Motivation: Fluorescent and luminescent reporter gene systems in combination with automated microplate readers allow real-time monitoring of gene expression on the population level at high precision and sampling density. This generates large amounts of data for the analysis of which computer tools are missing to date.

Results: We have developed WellReader, a MATLAB program for the analysis of fluorescent and luminescent reporter gene data. WellReader allows the user to load the output files of microplate readers, remove outliers, correct for background effects and smooth and fit the data. Moreover, it computes biologically relevant quantities from the measured signals, notably promoter activities and protein concentrations, and compares the resulting expression profiles of different genes under different conditions.

Availability: WellReader is available under a LGPL licence at <http://prabi1.inrialpes.fr/trac/wellreader>.

Contact: hidde.de-jong@inria.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received and revised on November 23, 2009; accepted on January 12, 2010

1 BACKGROUND

Experimental techniques based on fluorescent and luminescent reporter gene systems have become powerful tools for the real-time monitoring of gene expression in living cells. The reporter systems are obtained by fusing the promoter region (and possibly part of the coding region of a gene of interest) to a reporter gene. The expression of the reporter gene, which generates a light signal (fluorescence or luminescence) that is easy to measure, thus reflects the expression of the gene of interest [e.g. Zaslaver *et al.* (2006)].

Thermostated and agitated automated microplate readers allow emitted fluorescence and luminescence signals to be measured on the level of populations of microorganisms. They are typically designed to read a 96-well microplate, where each well contains a cell culture expressing a reporter gene or a control (cells without reporter constructs or growth medium only). All wells are successively read to measure in a single experiment the optical density of the cell culture, which is approximately proportional to the number of cells in the population, and the luminescence and fluorescence

intensity. This allows *in vivo* monitoring of the expression of several dozens of genes in parallel. Several applications in systems biology and biotechnology have confirmed the interest of monitoring gene expression by means of this approach [Kobiler *et al.* (2005); Lu *et al.* (2002); Onnis-Hayden *et al.* (2009); Ronen *et al.* (2002)].

In order to correctly interpret the reporter gene measurements, the primary data produced by the microplate reader must be translated into biologically relevant quantities, such as promoter activities and (relative) protein concentrations. This requires background corrections, data smoothing and fitting procedures, and mathematical transformations of the measured signals. Computer tools are essential for this, since automated microplate readers generate huge amounts of data, typically several thousands of data points per experiment. To our knowledge, no user-friendly computer tools for analyzing population-level fluorescence and luminescence reporter gene data exist in the public domain. The program WellReader aims at filling up this gap, thus facilitating the exploitation of the technology for the monitoring of gene expression in microorganisms.

2 DESCRIPTION OF FUNCTIONS

WellReader is equipped with a graphical user interface, structured around a visual representation of the microplate. This allows the user to access and control the analysis functions of the program (Fig. 1): an outlier detection module, a data smoothing and fitting module, a background correction module, modules for computing promoter activities and protein concentrations and a module for comparing expression profiles. In each step of the analysis, intermediate results can be stored for later use and exported to MATLAB for further analysis.

Outliers occasionally occur due to instrument or experimental errors, and they are removed either manually or by a simple outlier detection algorithm that can be parametrized by the user. After elimination of outliers, the data are interpolated by means of cubic smoothing splines with a user-defined smoothing parameter. Background levels of absorbance, fluorescence and luminescence systematic may cause systematic errors in the measurements. WellReader therefore allows the user to define background wells, and subtract the background levels from the uncorrected signals after appropriate normalizations. Relative measures of protein concentrations and promoter activities are then computed from the corrected absorbance, fluorescence and luminescence data.

*To whom correspondence should be addressed.

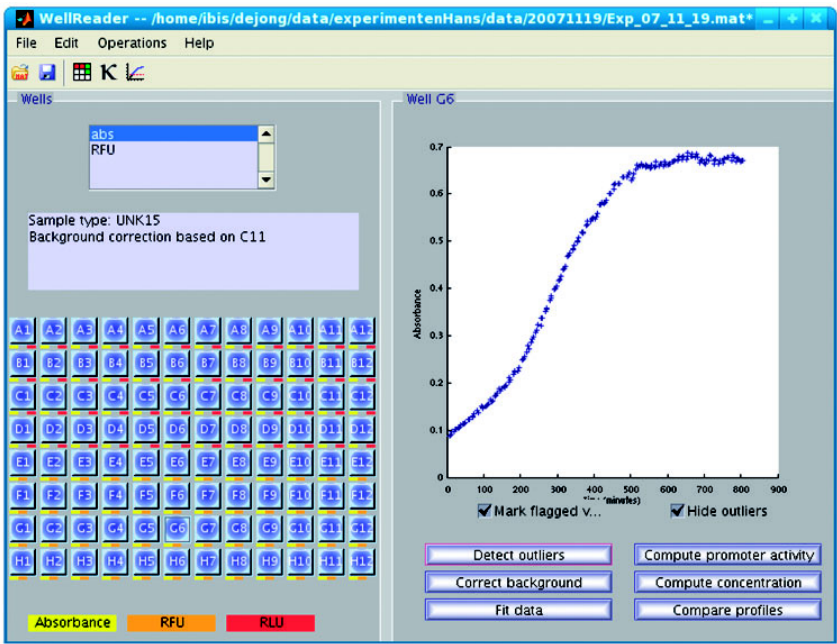


Fig. 1. Screenshot of WellReader main window. It shows the microplate and the absorbance data for one of the wells. The analysis functions listed at the lower right-hand side of the screen can be applied to the selected well.

This involves kinetic models of the temporal expression of reporter genes, as well as user-defined values for the half-lives of reporter proteins (e.g. GFP and luciferase) and host proteins. A detailed description of the data analysis procedures can be found in the Supplementary Material of this article.

All functions operate on the same underlying data structure, thus allowing the user to switch back and forth between the stages of the analysis process, and quickly investigate the consequences of, for example, different choices for smoothing parameters or background corrections.

3 IMPLEMENTATION AND DOCUMENTATION

WellReader has been implemented in MATLAB using the Spline toolbox (MathWorks). The program has been tested under Windows and Linux with different versions of MATLAB (R2007a, R2009a). In addition to the source code, a stand-alone executable is available to run WellReader outside the MATLAB environment. WellReader imports data in the plain text format generated by the Fusion Alpha microplate reader of Perkin Elmer as well as an XML format. This allows the user to load data from other plate readers, after a suitable transformation of the output files of the latter.

An online user manual is available at the WellReader web site together with a tutorial and an example dataset, concerning

key global regulators involved in different stress responses of *Escherichia coli*. The results of the analysis of the dataset will be reported elsewhere.

ACKNOWLEDGEMENT

The authors would like to thank Thomas Jouve for his contribution to a pilot version of the program.

Funding: European Commission (HYGEIA FP6-NEST-004995, COBIOS FP6-2005-NEST-PATH-COM/043379).

Conflicts of Interest: none declared.

REFERENCES

Kobiler, O. *et al.* (2005) Quantitative kinetic analysis of the bacteriophage λ genetic network. *Proc. Natl Acad. Sci. USA*, **102**, 4470–4475.
Lu, C. *et al.* (2002) Differential rates of gene expression monitored by green fluorescent protein. *Biotechnol. Bioengin.*, **79**, 429–437.
Onnis-Hayden, A. *et al.* (2009) Prokaryotic real-time gene expression profiling for toxicity assessment. *Environ. Sci. Technol.*, **43**, 4574–4581.
Ronen, M. *et al.* (2002) Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl Acad. Sci. USA*, **99**, 10555–10560.
Zaslaver, A. *et al.* (2006) A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat. Methods*, **3**, 623–628.

Supplemental Material of 'WellReader: A MATLAB Program for the Analysis of Fluorescence and Luminescence Reporter Gene Data'

Frédéric Boyer,¹ Bruno Besson,¹ Guillaume Baptist,^{1,2}
Jérôme Izard,^{1,2} Corinne Pinel,^{1,2} Delphine Ropers,¹
Johannes Geiselman,^{1,2} and Hidde de Jong¹

¹INRIA Grenoble – Rhône-Alpes

655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier Cedex, France

²Laboratoire Adaptation et Pathogénie des Microorganismes (CNRS UMR 5163)
Université Joseph Fourier, Bâtiment Jean Roget, Faculté de Médecine-Pharmacie
Domaine de la Merci, 38700 La Tronche, France

January 18, 2010

The Supplemental Material briefly summarizes the models and algorithms used in WellReader for the treatment of the measurements of absorbance, luminescence, and fluorescence intensities. In addition, we describe a data set obtained from the profiling of some key global regulators of *E. coli* during carbon starvation. For a user manual, tutorial, and download instructions for WellReader, consult the web site at <http://prabi1.inrialpes.fr/trac/wellreader>.

S1 Spline fitting

The absorbance, luminescence, and fluorescence data for a well are fitted by means of cubic smoothing splines, using the `csaps` function of the Spline toolbox in MATLAB (MathWorks). WellReader uses the default smoothing parameter of `csaps`, but the user can change this value through the graphical user interface. Outlier detection also uses spline fitting. The procedure removes the data points with the largest residuals, up to a user-defined threshold.

S2 Background correction

We denote by $A_u(t)$ the spline fit of the measured absorbance intensities over time, for a particular well. Idem $A_b(t)$ for the background absorbance, which is detected by performing measurements on wells without cells, containing growth medium only. The corrected absorbance $A(t)$ is defined as

$$A(t) = A_u(t) - A_b(t). \tag{S1}$$

The fluorescence background is usually determined by performing measurements on a strain carrying a reporter system with an inactive promoter. The fluorescence background activity is not constant, but rather varies with the population size due to the autofluorescence of cells. In this case, direct subtraction of the background readings from the uncorrected fluorescence intensity at each time-point t is not appropriate. We have to take into account that the size of the population generating the uncorrected signal may be different from the size of the population generating the background signal.

We therefore first compute the average fluorescence intensity per cell for the uncorrected signal and the background signal. We denote by $B(t)$ the absorbance of the strain carrying the inactive reporter system, $A(t)$ the absorbance of the strain with the functional reporter system, $I_u(t)$ the uncorrected fluorescence intensity, and $I_b(t)$ the background fluorescence intensity. The average fluorescence intensity per cell for the uncorrected and the background signal are then given by $I_u(t)/A(t)$ and $I_b(t)/B(t)$, respectively. We subtract the latter from the former to obtain the corrected average fluorescence intensity per cell, which we then multiply by the population size, as estimated by the absorbance, to obtain the corrected fluorescence intensity $I(t)$:

$$\begin{aligned} I(t) &= A(t) \left(\frac{I_u(t)}{A(t)} - \frac{I_b(t)}{B(t)} \right) \\ &= I_u(t) - \frac{A(t)}{B(t)} I_b(t). \end{aligned} \tag{S2}$$

The background correction for the luminescent measurements is in principle carried out in the same way as that for the fluorescence measurements. However, in practice the luminescence background is quite low, so that simple background subtraction is usually sufficient:

$$I(t) = I_u(t) - I_b(t). \tag{S3}$$

S3 Computation of promoter activities and protein concentrations

The computation of reporter protein concentrations and promoter activities is standard, following (with small modifications explained below) approaches described in [5, 6] (see also [4, 8] for the analysis of gene reporter data on the level of individual cells and cell populations).

S3.1 Concentration of reporter protein

The (corrected) fluorescence and luminescence levels are assumed proportional to the number of light-emitting molecules in the cells (*e.g.*, GFP and luciferase). Similarly, the absorbance level is proportional to the number of cells in a bacterial population. We thus infer that the cellular concentration $Y(t)$ of the reporter protein is proportional to the quotient of $I(t)$ and $A(t)$, that is,

$$Y(t) \sim \frac{I(t)}{A(t)}. \tag{S4}$$

In WellReader we define $Y(t)$ equal to $I(t)/A(t)$. Given that absorbance measurements are dimensionless, this amounts to expressing concentrations in terms of units RFU and RLU, the units in which fluorescence and luminescence intensities are reported. Notice that this results in concentration measures that are relative to a proportionality constant, as is usual in this kind of experiments. When this does not lead to ambiguities, we simply speak of concentrations instead of relative concentrations when we refer to variables with units RLU and RFU.

S3.2 Promoter activity

The promoter activity can be defined by means of the following kinetic model of the temporal expression of the reporter gene:

$$\frac{dY(t)}{dt} = F(t) - (\mu(t) + \gamma_y) Y(t), \quad (\text{S5})$$

where $F(t)$ represents the synthesis rate of the reporter protein, γ_y its degradation constant, and $\mu(t)$ the growth rate. In the absence of post-transcriptional regulation, $F(t)$ varies with the rate of transcription of the reporter gene, and is therefore called promoter activity. In the case of transcriptional fusions, the promoter activity of the reporter is a good indicator of the promoter activity of the host gene.

Rewriting (S5) gives the following expression for the promoter activity:

$$F(t) = \frac{dY(t)}{dt} + (\mu(t) + \gamma_y) Y(t). \quad (\text{S6})$$

Substituting (S4), and the following classical formula for the growth rate,

$$\mu(t) = \frac{d \ln A(t)}{dt} = \frac{dA(t)}{dt} \frac{1}{A(t)} \quad (\text{S7})$$

yields

$$F(t) = \frac{dI(t)}{dt} \frac{1}{A(t)} - \frac{dA(t)}{dt} \frac{I(t)}{A(t)^2} + \left(\frac{dA(t)}{dt} \frac{1}{A(t)} + \gamma_y \right) \frac{I(t)}{A(t)} \quad (\text{S8})$$

$$= \frac{dI(t)}{dt} \frac{1}{A(t)} + \gamma_y \frac{I(t)}{A(t)}. \quad (\text{S9})$$

The latter expression is used in WellReader for computing the promoter activity, given a user-specified value for the degradation constant γ_y . Notice that for $\mu(t) \gg \gamma_y$, we find

$$F(t) \approx \frac{dI(t)}{dt} \frac{1}{A(t)} \quad (\text{S10})$$

the expression for the promoter activity usually found in the literature (*e.g.*, [6]). More complex models can be used to analyze the data [4, 8], but they require an increased number of parameters and measurements on both the level of mRNAs and proteins. WellReader aims at giving a first assessment of the data which can be refined through further analysis, if needed, within the MATLAB environment.

S3.3 Correction for different half-lives of reporter and host proteins

In general, the stability of the reporter protein is not the same as that of the host protein, so that the concentration determined from (S4) is not necessarily representative of the concentration of the latter. Under certain conditions, reporter gene measurements can inform us about the concentration of the host protein too. In order to see this, we extend the kinetic model (S5) with an explicit equation for the temporal evolution of the concentration of the host protein, denoted by $X(t)$:

$$\frac{dX(t)}{dt} = F(t) - (\mu(t) + \gamma_x) X(t), \quad (\text{S11})$$

where γ_x is the degradation constant of the host protein. In the case of transcriptional fusions, the promoter activity $F(t)$ is the same as for the reporter protein, as explained above.

In order to solve (S11), the initial concentration $X(0)$ of the host protein needs to be known. Given that the degradation constants of the reporter and host proteins are generally different, the initial concentration of the host protein will be different from the initial concentration of the reporter ($X(0) \neq Y(0)$). It is possible to estimate the value of $X(0)$, if the system is in steady state before the start of the experiment, that is,

$$\frac{dY(0)}{dt} = \frac{dX(0)}{dt} = 0. \quad (\text{S12})$$

By means of (S11) and (S4), and the assumption that the growth rate is negligible at the beginning of the experiment, we obtain the following expression for the initial concentration of the host protein:

$$X(0) = \frac{\mu(0) + \gamma_y}{\mu(0) + \gamma_x} Y(0) \approx \frac{\gamma_y}{\gamma_x} Y(0) = \frac{\gamma_y}{\gamma_x} \frac{I(0)}{A(0)}. \quad (\text{S13})$$

Using this expression, we can express the host protein concentration in terms of the measured quantities, by solving the differential equation (S11) [1]:

$$X(t) = \exp^{-\int_0^t (\mu(\tau) + \gamma_x) d\tau} \left(\int_0^t F(\tau) \exp^{\int_0^\tau (\mu(\sigma) + \gamma_x) d\sigma} d\tau + X(0) \right). \quad (\text{S14})$$

We first infer by means of (S7) that $\int_0^t (\mu(\tau) + \gamma_x) d\tau = \ln(A(t)/A(0)) + \gamma_x t$. After substitution into (S14), and using the expression for the promoter activity (S9), we obtain

$$X(t) = \int_0^t \left(\frac{dI(\tau)}{d\tau} \frac{1}{A(t)} + \gamma_y \frac{I(\tau)}{A(t)} \right) \exp^{-\gamma_x(t-\tau)} d\tau + \frac{\gamma_y}{\gamma_x} \frac{I(0)}{A(t)} \exp^{-\gamma_x t}. \quad (\text{S15})$$

Integration by parts and rearranging the terms gives:

$$X(t) = \frac{I(t)}{A(t)} + \left(\frac{\gamma_y - \gamma_x}{\gamma_x} \right) \frac{I(0)}{A(t)} \exp^{-\gamma_x t} + (\gamma_y - \gamma_x) \int_0^t \frac{I(\tau)}{A(t)} \exp^{-\gamma_x(t-\tau)} d\tau. \quad (\text{S16})$$

Notice that if $\gamma_x = \gamma_y$, the expression simplifies to $X(t) = I(t)/A(t) = Y(t)$. That is, if the reporter and the host protein have the same half-life, the concentration of the reporter equals that of the host protein, as expected. The last two terms express the deviation of $X(t)$ from $Y(t)$ due to their difference in half-life. WellReader uses (S16) to compute the host protein concentration $X(t)$ from user-defined values for γ_y and γ_x .

	1	2	3	4	5	6	7	8	9	10	11	12
A	<i>rrnB</i>	<i>rpoSl</i>	<i>gyrA</i>	<i>topA</i>	<i>crpl</i>	<i>fis</i>	<i>rpoSc</i>	<i>gyrB</i>		BW25311	M9	lum bg
B	<i>rrnB</i>	<i>rpoSl</i>	<i>gyrA</i>	<i>topA</i>	<i>crpl</i>	<i>fis</i>	<i>rpoSc</i>	<i>gyrB</i>		BW25311	M9	lum bg
C	<i>rrnB</i>	<i>rpoSl</i>	<i>gyrA</i>	<i>topA</i>	<i>crpl</i>	<i>fis</i>	<i>rpoSc</i>	<i>gyrB</i>		BW25311	M9	lum bg
D	<i>rrnB</i>	<i>rpoSl</i>	<i>gyrA</i>	<i>topA</i>	<i>crpl</i>	<i>fis</i>	<i>rpoSc</i>	<i>gyrB</i>		BW25311	M9	lum bg
E	<i>rrnB</i>	<i>rpoSl</i>	<i>gyrA</i>	<i>topA</i>	<i>crpl</i>	<i>fis</i>	<i>rpoSc</i>	<i>gyrB</i>	<i>crpc</i>	BW25311	M9	fluo bg
F	<i>rrnB</i>	<i>rpoSl</i>	<i>gyrA</i>	<i>topA</i>	<i>crpl</i>	<i>fis</i>	<i>rpoSc</i>	<i>gyrB</i>	<i>crpc</i>	BW25311	M9	fluo bg
G	<i>rrnB</i>	<i>rpoSl</i>	<i>gyrA</i>	<i>topA</i>	<i>crpl</i>	<i>fis</i>	<i>rpoSc</i>	<i>gyrB</i>	<i>crpc</i>	BW25311	M9	fluo bg
H	<i>rrnB</i>	<i>rpoSl</i>	<i>gyrA</i>	<i>topA</i>	<i>crpl</i>	<i>fis</i>	<i>rpoSc</i>	<i>gyrB</i>	<i>crpc</i>	BW25311	M9	fluo bg

Table S1: Microplate layout for an experiment in which the expression of some global regulators in *E. coli* are monitored. The name of the gene of which the promoter region is cloned into the reporter plasmid is shown in italic (the two variants of *crp* and *rpoS* are postfixed with the characters *c* and *l*). The columns labelled BW25311, M9, lum bg, and fluo bg refer to wells containing bacteria without reporter plasmids, growth medium only, bacteria carrying a luminescent reporter system with an inactive promoter, and bacteria carrying a fluorescent reporter system with an inactive promoter, respectively. The wells in rows A-D contain luminescent reporter plasmids and those in E-H fluorescent reporter plasmids.

S4 Expression monitoring of global regulators in *E. coli*

Global regulators form the backbone of the network enabling the bacterium *E. coli* to adapt its functioning to environmental stress conditions [7]. We measured the expression of the genes encoding some of these global regulators by means of fluorescent and luminescent reporters. The regulators involved are Fis, Crp, GyrA, GyrB, TopA, and RpoS. Moreover, we monitored the expression level of the operons coding for the stable RNAs. The promoter regions of the corresponding genes were cloned into *gfp*- and *lux*-containing plasmids that are derivatives of plasmids pZE1RM [3] and pSB377 [2], respectively. For *crp* and *rpoS* we constructed two different reporter plasmids, in order to monitor expression from all or some of the known promoters of these genes.

Glycerol stocks, stored at -80°C, of strains BW25311 carrying a plasmid-encoded reporter gene were grown overnight at 37°C, with shaking at 200 rpm, in M9 minimal medium supplemented with 0.45% glucose. For plasmid-carrying strains, the growth medium was supplemented with 100 $\mu\text{g}\cdot\text{ml}^{-1}$ ampicillin. At the start of the experiment, the batch cultures grown overnight were diluted between 10 and 50 times into the same minimal medium with glucose and grown in a microplate for about 12 h at 37°C. The microplate was agitated at regular intervals during growth in the Fusion microplate reader (Perkin-Elmer) and a glass bead was added to the wells to improve stirring. During a typical experimental run we acquired about 100 readings each of absorbance, luminescence, and fluorescence. Fluorescence excitation was at 485 nm and emission was monitored at 520 nm. Absorbance measurements used a 600 nm filter. The microplate layout is shown in Table. S1.

Both the raw data file (.CVS) and the WellReader file (.mat) are available at the WellReader website. The results of the analysis of the data from this experiment will be reported elsewhere.

References

- [1] R. Courant. *Differential and Integral Calculus*, 2 vols. Blackie & Son, London, 1959.

- [2] S. Déthiollaz, P. Eichenberger, and J. Geiselmann. Influence of DNA geometry on transcriptional activation in *Escherichia coli*. *EMBO J.*, 15(19):5449–5458, 1996.
- [3] M.B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 2000.
- [4] B. Finkenstädt, E.A. Heron, M. Komorowski, K. Edwards, S. Tang, C.V. Harper, J.R.E. Davis, M.R.H. White, A.J. Millar, and D.A. Rand. Reconstruction of transcriptional dynamics from gene reporter data using differential equations. *Bioinformatics*, 24(24):2901–2907, 2008.
- [5] J.H.J. Leveau and S.E. Lindow. Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. *J. Bacteriol.*, 183(23):6752–6762, 2001.
- [6] M. Ronen *et al.* Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl. Acad. Sci. USA*, 99(16):10555–10560, 2002.
- [7] D. Ropers, H. de Jong, M. Page, D. Schneider, and J. Geiselmann. Qualitative simulation of the carbon starvation response in *Escherichia coli*. *Biosystems*, 84(2):124–152, 2006.
- [8] X. Wang, B. Errede, and T.C. Elston. Mathematical analysis and quantification of fluorescent proteins as transcriptional reporters. *Biophys. J.*, 94(6):2017–2026, 2008.

3.3 Identification de nouveaux régulateurs transcriptionnels

Une des applications possibles pour cette méthode d'acquisition et d'analyse de données, est la mise en évidence de régulateurs d'un gène donné. Par exemple, dans le cadre de la thèse de Guillaume Baptist, nous avons réalisé un crible de l'ensemble des mutants de la collection Keio (Baba *et al.*, 2006), afin d'identifier des régulateurs du gène *acs*, qui permet l'utilisation de l'acétate comme source de carbone. Cette collection recense l'ensemble des simples mutants, non-essentiels, d'*Escherichia coli*, soit 3911 gènes délétés. En comparant la dynamique d'expression d'*acs* dans les mutants avec le profil observé dans la souche sauvage, nous avons pu identifier des gènes impliqués dans la régulation transcriptionnelle de ce gène. Parmi ces gènes, nous avons retrouvé des régulateurs déjà connus, tels que le complexe CRP-AMPc. En effet, l'expression d'*acs* était complètement abolie dans un mutant *crp* ainsi que dans le mutant *cyaA*, gène codant pour l'adénylate cyclase, l'enzyme responsable de la production d'AMPc. Nous avons également pu identifier de nouveaux régulateurs, impliqués dans le métabolisme central, comme PtsG, qui participe à l'import du glucose dans la cellule, ou encore Pgi et PfkA, deux enzymes impliquées dans la glycolyse. Dans ces différents mutants, l'expression d'*acs* était également abolie ou fortement diminuée. Cette étude fait l'objet d'une publication que vous trouverez en annexe de cette thèse. Au cours de ma thèse, j'ai plus particulièrement utilisé cette approche pour caractériser la dynamique d'expression du gène *crl* et identifier de nouveaux régulateurs de sa transcription.

3.4 Étude de la régulation du gène *crl*

La protéine Crl a tout d'abord été identifiée comme un régulateur de l'expression des curli, des structures de surface qui permettent l'adhésion sur différents supports, biotiques ou non. En effet, Crl active l'expression de l'opéron *csgBA*, qui code les différentes protéines composant et exportant les curlis (Arnqvist *et al.*, 1992). Néanmoins, Crl ne possède pas de domaine de liaison à l'ADN, et exerce son activation par l'intermédiaire du facteur sigma RpoS avec lequel il interagit directement (Pratt & Silhavy, 1998 ; Bougdour *et al.*,

2004; England *et al.*, 2008). Crl paraît donc plutôt être un modulateur de l'activité de σ^S , qui est un des régulateurs majeurs d'*Escherichia coli*.

Le facteur sigma alternatif RpoS est l'acteur principal de la réponse au stress. Lorsque la cellule est soumise à différents stress et notamment lors d'une carence nutritionnelle, le facteur σ^S s'accumule et dirige l'ARN polymérase vers la transcription de gènes d'adaptation et de résistance. L'ensemble des gènes transcrits par l'holoenzyme $E\sigma^S$ constitue son régulon. Ce dernier a été particulièrement étudié au cours des dernières années (Lacour & Landini, 2004; Patten *et al.*, 2004; Weber *et al.*, 2005; Dong *et al.*, 2008; Dong & Schellhorn, 2009), et représente près de 10% du génome d'*E. coli*.

La régulation de RpoS est complexe et implique de nombreux mécanismes jouant autant au niveau transcriptionnel qu'au niveau de la stabilité de la protéine (Klauck *et al.*, 2007). Parmi les régulateurs, on peut citer l'anti-facteur σ^{70} : Rsd, le nucléotide signal : ppGpp ou encore la protéine DksA, qui vont influencer la compétition entre les différents sigma en défavorisant σ^{70} au profit des facteurs alternatifs (Jishage *et al.*, 2002; Paul *et al.*, 2004; Gourse *et al.*, 2006; Bougdour & Gottesman, 2007). Crl joue également un rôle dans la régulation de RpoS, en augmentant l'affinité de σ^S pour l'enzyme coeur de l'ARN polymérase et en favorisant l'initiation de la transcription par RpoS : liaison et ouverture des promoteurs RpoS-dépendants (Bougdour *et al.*, 2004; Gaal *et al.*, 2006; Robbe-Saule *et al.*, 2006; England *et al.*, 2008; Dudin *et al.*, en préparation).

Ce rôle de modulateur de l'activité de RpoS, et le fait qu'il régule avec lui l'expression d'un très grand nombre de gènes, démontre que Crl est un régulateur important d'*Escherichia coli*. Malgré sa forte implication dans le réseau de régulation de cette bactérie, peu de choses sont connues concernant la régulation de son expression au cours de la croissance. Nous nous sommes donc attardés à mieux caractériser son promoteur, sa dynamique d'expression et à en chercher les régulateurs.

Ces travaux font l'objet d'un article en préparation (ci-joint), intitulé « CRP and Fis regulate the expression of the Crl regulatory protein in *Escherichia coli* », dont je vais présenter les principaux résultats.

Dans un premier temps, nous avons construit un plasmide rapporteur portant la région promotrice de *crl* (Figure 3.1) et mesuré son expression tout au long d'une croissance sur milieu minimum et milieu riche. Nous avons ainsi montré que l'expression de *crl* présentait

un pic transitoire lors de la phase exponentielle (Fig. 1 de l'article). Des gels de retard réalisés avec de l'ARN polymérase contenant σ^{70} ou σ^S , nous ont permis de montrer que le promoteur de *crl* était essentiellement reconnu par le facteur σ^{70} (Fig. 2 A de l'article). Ce résultat a été confirmé par une empreinte au permanganate de potassium, montrant une ouverture préférentielle du promoteur par l'holoenzyme contenant σ^{70} (Fig. 2 B et C de l'article). Ceci est cohérent avec l'expression de *crl* qui est majoritairement en phase exponentielle et précède l'accumulation de la protéine RpoS, avant l'entrée en phase stationnaire de croissance (Robbe-Saule *et al.*, 2006).

Nous avons également trouvé que le pic d'expression de *crl* était indépendant de l'entrée en phase stationnaire. Ce qui suggère une régulation de ce gène par une molécule signal ou par des régulateurs globaux dont l'expression est aussi maximale lors de la phase exponentielle, notamment la nucléoprotéine Fis (Azam *et al.*, 1999). Afin d'identifier les régulateurs responsables de ce pic d'expression, nous avons mesuré l'activité de *crl* dans une centaine de mutants de la collection Keio (Baba *et al.*, 2006). Cette sélection comporte des gènes impliqués dans le Quorum Sensing, codant pour des systèmes à deux composants ainsi que des nucléoprotéines et des régulateurs globaux. Ce crible a mis en évidence deux mutants, *fis* et *crp*, pour lesquels l'expression de notre plasmide rapporteur était grandement affectée (Fig. 2 et 3 de l'article).

Lorsque le gène *fis* est absent, l'expression de *crl* est diminuée d'environ 2 fois, ce qui suggère que la nucléoprotéine Fis active la transcription de *crl*. Nous avons confirmé *in vitro* que la protéine Fis se fixait sur la région promotrice de *crl* suggérant un mécanisme d'activation direct de son promoteur (Fig. 5 A et Fig. 6 de l'article). Concernant le complexe CRP-AMPc, nous avons montré *in vivo* qu'il réprimait l'expression de *crl* puisque lorsque l'on mute le gène *crp* ou le gène *cyaA*, qui code l'adénylate cyclase (l'enzyme qui produit l'AMPc), l'expression de *crl* augmente. Néanmoins, cette répression est probablement indirecte puisque le complexe CRP-AMPc ne semble pas se fixer sur le promoteur de *crl* (Fig. 5 B de l'article).

Publication n°2 :

CRP and Fis regulate the expression of the Crl
regulatory protein in *Escherichia coli*

**CRP and Fis regulate the expression of the Crl regulatory protein in
*Escherichia coli***

Jérôme Izard¹ *et al.*

¹ Laboratoire Adaptation et Pathogénie des Micro-organismes, Institut Jean Roget,
Domaine de la merci, Université Joseph Fourier, BP 170, 38042 Grenoble cedex 9, France

#To whom correspondence should be addressed:

stephan.lacour@ujf-grenoble.fr

Tel: 0033 476 63 74 84

Fax: 0033 476 63 74 97

Running title: CRP and Fis regulate the *crl* expression in *E. coli*

ABSTRACT

In *Escherichia coli*, the sigma factor RpoS controls the expression of about 500 genes in response to a variety of stress conditions. The Crl protein enhances RpoS activity in a promoter specific fashion, but little is known regarding its own transcriptional regulation. We show here that Crl is transcribed from a single σ^{70} -dependent promoter. During a batch culture in minimal medium, the *crl* promoter is induced in a transitory fashion with a maximal activity in mid-exponential phase. Two global regulators, Fis and CRP, were respectively found to activate or repress the transcription of *crl*. A molecular characterization shows that Fis binds to the *crl* promoter and directly activates transcription of *crl*, whereas CRP represses *crl* transcription by an indirect mechanism.

INTRODUCTION

To deal with environmental changes, such as nutrient limitation, osmotic or oxidative stress, *Escherichia coli* induces expression of stress response genes. The main actor of this response is a sigma subunit of RNA polymerase (RNAP), σ^S (RpoS). In bacteria, each σ subunit controls transcription initiation at a specific set of promoters with common DNA features. During fast growth, most of the essential genes are transcribed by the holoenzyme containing the σ^{70} subunit. Upon stress condition, the accumulation of the alternative sigma S factor can strongly rewire gene expression. Thus, RpoS addresses the RNA polymerase to a new set of promoters and stimulates the transcription of stress response genes or others necessary for stationary phase survival. Its extensive regulon had been characterized by several gene arrays in various growth or stressfull conditions and exceeds 12% of the *E. coli* genome (1–6). Since RpoS was found to regulate various metabolic genes, it is considered as the second vegetative sigma factor during stationary growth phase in addition of being a key regulator of adaptive responses. These observations are comforted by the extremely complex network which modulates RpoS accumulation and activity from transcription of *rpoS* to the protein stability (7, 8). For instance during exponential growth, RpoS is subjected to an active degradation by the ClpXP protease due to a specific addressing by the adaptator protein RssB (9). Later during the growth in stationary phase, its activity is indirectly enhanced by the reduction of σ^{70} availability or effectiveness. Different actors are involved including at least a protein, an RNA and a signaling molecules. For instance, the alarmone ppGpp together with DksA, favors the association of the core RNAP with alternative sigma factors such as RpoS (Jishage et al., 2002; Paul et al., 2004; Gourse et al., 2006). The anti-sigma 70 factor, Rsd, sequesters free σ^{70} (10, 11) and therefore helps transcription by other sigma factors. The small RNA 6S which sequesters a large amount of σ^{70} by mimicking a promoter region also contributes to this rewiring of gene expression in stationary phase (12, 13).

Another protein factor, unable to bind DNA, was found to positively modulate RpoS activity. Initially identified as a positive modulator of the *E. coli* adhesion fimbriae called curli (14), the Crl protein appeared to enhance transcription of RpoS-dependent genes (15). Later, it was observed that Crl increases curli production *in E.*

coli and *Salmonella* because it stimulates promoter binding and opening by $E\sigma^S$ holoenzyme at the *csgB* and *csgD* promoters, i.e. of genes encoding the curli monomers or regulator (16, 17). Recent studies show that Crl directly modulates RpoS activity by either favoring $E\sigma^S$ holoenzyme formation or increasing promoter recognition and transcription initiation in a promoter dependent manner (17–20). Indeed, Western blots experiments show that Crl accumulates at the onset of stationary phase, slightly earlier than RpoS (16, 21). Furthermore, Crl generally increases the affinity of σ^S for the core RNA polymerase (20) but it might also contribute to the regulation of RpoS accumulation and it help this sigma factor to bind the core RNAP at low concentration (17, 22). In *Salmonella*, Crl affects the expression of RpoS-regulated genes at low temperature (below 30°C) where it seems to better accumulate in *Salmonella* (21) and several *E. coli* strains (5, 16, 22).

Despite multiple evidences that Crl positively modulates RpoS activity and the high overlapping between the Crl and RpoS regulons (22, 23), very little is known about the physiological conditions that modulate Crl expression. In this study, we focused on the control of *crl* transcription: we confirmed the effectiveness of the predicted promoter (14) and looked for transcriptional modulators of *crl*. For this purpose, we have monitored its expression profile over time during growth in current media and we observed a transitory induction of *crl* in exponential phase. Using a collection of *E. coli* mutant strains (24), we screened for genetic factors able to affect such a temporal expression pattern of *crl*. The disruption of the genes encoding the transcriptional regulators Fis and Crp was able to severely affect *crl* transcription. Fis was found to directly activate transcription of *crl* by binding upstream of the core promoter region.

MATERIAL & METHODS

Strains and plasmids

The strains used in this study are *Escherichia coli* K-12 strain BW25113 and isogenic insertion mutants of the Keio collection were used (24). The collection contains all single-gene knock-out of non essential genes of a W3110 derivative strain. The wild type (BW25113) and its *crp* or *fis* mutant strains were transformed with a reporter plasmid carrying the *crl* promoter fused to the *luxABCDE* operon as a reporter system (25). This vector is a low copy number plasmid with *colE1* replicon and an ampicillin resistance gene. To produce the transcriptional fusion, the sequence upstream of the *crl* coding region was cloned as a 370 bp XhoI-EcoRI fragment generated by PCR using genomic DNA of strain W3110 as template and the primers *crl_10* (5'-aaactcgagttaagtgccagcaacaggtg-3'), and *crl_13* (5'-aaagaattcctgttgatgcaactgttt-3'), annealing respectively from positions -360 to +10 relative to the transcription start site.

Growth conditions and reporter gene measurements

Cultures were grown either in M9 minimal medium supplemented with 0.2% glucose or in LB, always at 37° C. To maintain the reporter plasmid and mutations, we add into the medium respectively 100µg/ml of ampicillin and 50µg/ml of kanamycin.

Time-series measurements were carried out in 96 wells microplates, inoculated from a fresh overnight culture about 18h, in the very same medium used during measurements. Before running time-series the pre-culture were diluted 10 fold (if not otherwise specified) into 150µl of fresh medium. Then bacteria were grown in the thermostated microplate reader (Perkin Elmer, Fusion alpha) and shaken every ten min to insure micro-aerobic conditions and to prevent cell sedimentation.

Absorbance, at 600nm, and luminescence were measured for each well alternatively and roughly every 10 minutes during each time-series. The raw data are analysed using the Wellreader software (26) to compute the reporter gene concentration. Absorbance and luminescence measurements were background corrected as described in de Jong *et al*, (2010), to consider the growth medium absorbance and basal light - barely detectable - emitted by a strain carrying a

promoter-less reporter plasmid. Reporter gene concentration is the ratio of both the corrected luminescence and absorbance, and therefore is the light emitted per bacteria.

***In vitro* experiments**

For gel mobility shift assays, DNase I protection experiments and KMnO₄-reactivity assays, we used linear DNA fragments and followed the general procedure described in Lacour & Landini (2004) excepted that the core RNAP polymerase was purchased to Epicentre biotechnologies. Purified sigma factors as well as the CRP and Fis proteins were kindly provided by A. Kolb and D. Schneider. Labeled DNA was generated by PCR after 5'-phosphorylation of the primers Crl_C1 (5'-gtggagctgaatcgctat-3') or Crl_pro2 (5'-acgcacttccggtgccggtttcac-3') to produce respectively P32-labeled template or non-template strands. The two primer combinations crl_C1 / crl_C2 (5'-ctgctcttcgggtgtc-3') and Crl_pro2 / crl_11 (5'-aaactcgagccccatttaacccggtgtatc-3') respectively generate 328 pb and 300 pb overlapping fragments covering the small *yafA/crl* intergenic region (58 bp) which contains *Pcrl*.

Briefly, for gel mobility shift assays (GMSA), the reconstituted RNA polymerase (17 to 100 nM)–DNA (1 nM) complexes were allowed to form for 15 min at 37°C in K-glu100 buffer (40 mM HEPES, pH 8.0, 10 mM magnesium chloride, 100 mM potassium glutamate, 4 mM dithiothreitol, and 500 µg/ml bovine serum albumin), in a final reaction volume of 10 µl. The reaction mixture was then loaded onto a native 5% polyacrylamide gel after addition of 2.5 µl of loading buffer (50% sucrose with trace of xylene cyanol and bromophenol blues plus 150 µg/ml heparin). The same condition was used for GMSA with Fis (25-500 nM) both at 19°C and 37°C. To address CRP (100-500 nM) binding to *Pcrl*, cAMP was added to the K-glu100 buffer at 200µM and within the acrylamide gel (5%) or the TBE running buffer (0.5x). All the reactions were assayed at least twice and an additional assay was performed in KCl buffer, which gave the exact same results.

For DNaseI footprinting, Fis (50 to 200 nM) was incubated with both DNA fragments (3 nM) for 20 min at 37°C in K-glu100 buffer. The protein free DNA samples were treated with 1 µg/ml DNaseI for 20 s, while the incubation was prolonged to 30 s for protein–DNA complexes. Samples were phenol extracted,

precipitated and resuspended in formamide loading blue before being loaded onto a 7% polyacrylamide denaturing gel. KMnO₄ reactivity experiments were also performed in K-glu100 (without DTT) using 50 nM of either forms of RNA polymerase ($E\sigma^S$ and $E\sigma^{70}$) after incubation with the labeled promoter DNA for 15 min at 37°C. KMnO₄ was added to a final concentration of 10 mM and the reaction was stopped after 30 s by adding 2 mM DTT. Samples were then precipitated and loaded like the DNaseI samples onto a sequencing gel.

RESULTS

Expression profile of *crl* during growth in batch culture

In order to study the regulation of *crl* transcription, we have constructed a transcriptional fusion of the *crl* promoter (*Pcrl*) to the bacterial luciferase operon, *luxCDABE*. This construct allows us to measure the temporal expression profile of *crl* which is an important parameter to understand how Crl regulates RpoS activity depending on the growth stage. The Figure 1 A shows that *crl* is expressed in a transitory fashion in mid-exponential phase. This is consistent with our previous results showing Crl start to accumulate prior to the entry in stationary phase (16). However, the transcriptional activity reaches a maximal value after 120 min of growth independently of the time when cells enter stationary phase delayed by the addition of glucose (Fig. 1 A). As shown in Figure 1 B, if the batch culture is started with a more concentrated inoculum, e.g. 1/5 instead of 1/20, the peak of *crl* expression occurs earlier but roughly at the same absorbance of the resulting culture (about 0.2-0.3 at 600nm). Indeed, the magnitude of induction is not affected by the initial cell density. Therefore, Crl could be induced by a molecule which accumulates in the medium independently of the entry into stationary phase. A very strong induction is also observed in LB medium but with a shorter duration very early in the kinetic monitoring (data not shown).

The *crl* gene is transcribed from a σ^{70} promoter

In order to better understand the regulation of *crl* transcription, we characterized the promoter of *crl* using gel retardation and permanganate footprint and primer extension. The maximal activity of the *crl* promoter was identified during exponential phase which suggests a preferential transcription by the holoenzyme $E\sigma^{70}$. In addition, since Crl modulates RpoS activity, we tested whether $E\sigma^S$ could also transcribe *crl* to support its expression in stationary phase. The Figure 2 shows that the $E\sigma^S$ holoenzyme does not strongly bind the *crl* promoter to form Heparin-resistant complexes. In the same conditions, the $E\sigma^S$ holoenzyme forms stable complexes at low RNAP concentration at a *bona fide* RpoS-dependent promoter, *PaidB* which is poorly transcribed by $E\sigma^{70}$ (27, 28). In contrast, $E\sigma^{70}$ forms much more stable complexes at *Pcrl* than at *PaidB*. Thus, this band-shift assay clearly shows that $E\sigma^{70}$ better recognize and open the *crl* promoter as compared to $E\sigma^S$. This is consistent

with the absence of reduction of *crl* expression in an *rpoS* mutant strain, especially in stationary phase (data not shown).

The permanganate footprint confirms that only the $E\sigma^{70}$ holoenzyme is able to efficiently decatenate the *crl* promoter around its -10 hexamer. (Fig. 2 B). Four thymidine of the non-template strand react with $KMnO_4$ from positions -7 to -11 according to the transcription start site (TSS) of *crl* proposed (14). We performed a novel primer-extension and a Northern-blot analysis and both experimental approaches points to a single monocistronic mRNA (data not shown). We thus validate the functionality of the single promoter 19 bp upstream of the start codon determined by Arnqvist and co-authors. This promoter is transcribed from a single $E\sigma^{70}$ -dependent promoter. As expected from the *in vivo* expression data, the $E\sigma^S$ holoenzyme does not open the *crl* promoter in our conditions.

Identification of *Pcrl* regulators

The fact that *crl* presents a peak of transcription always at the same growth state (Fig. 1) suggests that its transcription could be regulated by a molecule that accumulates in the medium during the growth of the bacterial culture. Therefore, we have measured the transcriptional activity of *Pcrl*, using the same luciferase reporter construct, in a selection of *E. coli* mutants (24) involved in environmental sensing. We tested 11 mutants related to the quorum sensing (QS) including biosynthetic genes for autoinducer (AI) production. We also selected 64 mutants of two-component systems (2CS) identified in *E. coli* (Table S1) as they sense and induce responses to a variety of environmental changes. Unfortunately, none of these mutants was able to affect the profile of *crl* induction in exponential phase neither in minimal nor in LB medium.

Neither QS nor 2CS are responsible of *crl* induction before entry into stationary phase. Another hypothesis to explain the transitory peak of *crl* activity relies on an internal control of the transcription machinery by common transcription factors such as nucleoid associated proteins (NAPS). The cellular content in nucleoproteins is indeed drastically changing during growth transition from exponential to stationary phase (29). Since *rpoS* transcription is regulated by Fis, we tested if mutations in gene encoding NAPS also affect *crl* expression. Glucose exhaustion results in modification of the cellular content in signaling molecules such as cAMP and ppGpp.

Thus, we also tested mutants unable to produce these effectors. Among the 16 mutants tested (Table S1) we found two mutations in genes encoding for global regulators able to strongly affect the expression of *crl*: CRP and Fis. Therefore, we focused on the modes of action of these regulators to control *crl* transcription.

Fis activates *Pcrl*

In *E. coli*, the nucleoid composition is sensitive to the growth rate and Fis represents the major nucleoprotein during exponential phase (29). We tested if strong Fis accumulation in mid exponential phase could explain *crl* induction. As expected, the *fis* deletion decreases *crl* expression during active cell division as compared to expression in the wild-type strain (Fig. 3). Indeed, the *fis* mutation barely affects *crl* transcription during stationary phase. Consistently, Fis overexpression slightly increases *crl* expression (data not shown) albeit the fact that Fis overproduction seriously affects *E. coli* cell growth (Ranquet *et al*, article in preparation).

Since our *in vivo* expression data suggest that Fis activates *crl* expression, we tested the interaction of the purified protein with *Pcrl*. The gel-retardation assay clearly shows that Fis is able to bind to the *crl* promoter with an affinity closed to 25nM (Fig. 5 A). The DNaseI footprint experiment confirms such a strong interaction which results both in multiple protections and hypersensitivities (Fig. 6). The Fis binding site seems to be located upstream of the -35 promoter, between -30 and -80 positions according to the TSS. This location is consistent with a transcriptional activation since it allows interactions with either sigma or alpha subunits of RNAP. It is also compatible with the loss of effect of the *fis* mutation *in vivo* on transcription from a shortened *Pcrl* including only 40 nucleotides upstream of the TSS (data not shown). Interactions are also observed within the coding sequence near the start codon on the template strand which suggests wider interactions of Fis with *Pcrl* in the absence of RNAP. Our results tend to prove a direct activation of the *crl* promoter by Fis.

The CRP-cAMP complex represses the transcription of *Pcrl*

Our selective screening of mutant indicates another global regulator, CRP, capable of modifying *crl* expression profile. In a *crp* mutant strain, the peak of *crl* expression is higher than in a wild type strain and the mutation also increases the duration of *crl* induction during the whole course of exponential phase (Fig. 4). Since CRP needs cAMP to bind DNA, we transformed our reporter fusion into a *cyaA* (adenylate cyclase) mutant strain unable to produce cAMP. In this mutant, the transcription of *crl* was also stimulated and the expression profile was very similar to the pattern observed in the *crp* mutant strain (Fig. 4). Addition of cAMP (1mM) in the M9 medium results in a reduction of *Pcrl* activity during the whole experiment. These results are consistent and show that the cAMP-CRP complex acts as a repressor of *crl* transcription.

To determine whether this repression results from a direct binding of CRP to the *crl* promoter, we performed a gel shift assay in presence of cAMP (Fig. 5 B). CRP was not able to bind *Pcrl* in our conditions which yet allow binding of CRP to the *osmY* promoter, our positive control (30). This result suggests an indirect repression of *crl* transcription by the CRP-cAMP complex.

DISCUSSION

In this study, we characterized the expression profile of *crl* by using transcriptional fusions to the luciferase reporter gene. It allows us to measure the whole dynamic of transcription after a glucose up-shift in minimal medium or a fresh dilution in LB. We observed a peak of *crl* transcription during the first half of the exponential phase. We also observed such a transcriptional profile for different main regulators of *E. coli* using the GFP or luciferase as a reporter proteins (Ropers *et al*, in preparation). Thus, we decided to investigate the molecular basis of this typical induction pattern with *crl* as a new model of global regulator as it regulates RpoS activity. Interestingly, the cellular growth rate controls RpoS accumulation (31) and *rpoS* transcription is modulated by both CRP and Fis (32).

We have discovered that the nucleoprotein Fis activates *crl* transcription and directly binds *Pcrl* immediately upstream of the core promoter. Since Fis strongly accumulates during exponential phase to reach its maximal cellular concentration (29), it probably explains why *crl* transcription is also maximal during logarithmic growth. The location of Fis binding site at *Pcrl* is consistent with a direct activation as previously demonstrated for the activation of number of promoters where Fis binds at very different location from positions -41 to -120, e.g. at the *rmBP1* or the *tyrT* promoters (33, 34). Depending on such a location, Fis interacts with α -CTD in different ways to stimulate transcription (as a class I or II activator) and can even *cis*-act from a very distant site by translocation of superhelical energy, e.g. at the *leuV* promoter (35). Fis modulates the expression of genes with adaptive and protective functions (36) and was found to activate the *proP* P2 RpoS-dependent promoter (37, 38). In this way, it makes sense that Fis activates *crl* expression to reinforce adaptation to stationary phase.

We have also found that the CRP protein indirectly represses *crl* transcription. CRP also negatively regulates *rpoS* transcription (39) and different gene-chip experiments demonstrate that RpoS regulates a variety of metabolic genes including carbohydrate metabolism (2, 28) sometimes with the help of Crl (4). Thus, RpoS and Crl are not only key regulators of stress-responses but also global transcription factors whose expression is modulated by the catabolic repressor CRP.

The maximal level of *crl* transcription is always observed at barely the same cell density (Fig. 1) and mutation of neither *crp* nor *fis* completely abolishes *crl* induction. So, we have tried to identify a gene responsible for the production of a diffusible molecule able to induce *crl* in addition of Fis accumulation. None of the mutants of genes involved in quorum sensing did affect *crl* transcription. Beside well-characterized AI pathways, we found back a stimulatory effect of indole (as in Lelong *et al.*, 2007, but with a transcriptional fusion depleted of coding sequence) and of its precursor, the tryptophan amino acid (data not shown). Interestingly, two proteins involved in indole production, namely WrbA and TnaA, are positively regulated by RpoS (28) which suggests a positive feed-forward loop for RpoS activity. We also asked if the increased *P_{crl}* activity in exponential phase was triggered by one of the 30 two-component systems which integrate environmental signals and regulate gene expression. None of them were able to affect *crl* expression, despite the fact that several 2CS do modulate *rpoS* transcription (40).

Our study reveals that *crl* transcription is maximal during exponential phase right before RpoS accumulation (16, 21). This is compatible with the hypothesis that Crl helps the alternative sigma factor to transcribe adaptive genes prior to the entry into stationary phase, when RpoS is not yet detectable by western blot. Indeed, Crl was found to strongly stimulate RpoS-dependent transcription in a *Salmonella* strain producing low concentrations of RpoS (19). Therefore, it is consistent to find a common regulation of their expression by global regulators such as Fis and CRP. The strong accumulation of Fis in exponential phase might explain the precocious *crl* induction even though the *fis* mutation does not completely abolish *crl* expression. CRP also contributes to turn off *crl* expression especially in decelerating growth when cAMP accumulates before entry into stationary phase. Surprisingly, we were able to observe a succession of induction peaks after successive dilution of the culture (Fig. S1). This suggests that a secreted molecule could repress *crl* transcription when it reaches a critical concentration in the medium. This signal could be produced by a yet uncharacterized genetic factor to be identified with new blind screening. We are currently investigating the consequences of a decreased activity of the transcription machinery to explain the general induction profile observed for the transcription of global *E. coli* regulators. Our model might be very useful to explain the growth rate regulation of global regulators like Fis (41), RpoS (31) and Crl (this work).

REFERENCES

1. **Lacour S, Landini P.** 2004. SigmaS-dependent gene expression at the onset of stationary phase in *Escherichia coli*: function of sigmaS-dependent genes and identification of their promoter sequences. *Journal of bacteriology* **186**:7186–95.
2. **Patten CL, Kirchhof MG, Schertzberg MR, Morton RA, Schellhorn HE.** 2004. Microarray analysis of RpoS-mediated gene expression in *Escherichia coli* K-12. *Molecular genetics and genomics* : MGG **272**:580–91.
3. **Weber H, Polen T, Heuveling J, Wendisch VF, Hengge R.** 2005. Genome-wide analysis of the general stress response network in *Escherichia coli*: sigmaS-dependent genes, promoters, and sigma factor selectivity. *Journal of bacteriology* **187**:1591–603.
4. **Typas A, Barembruch C, Possling A, Hengge R.** 2007. Stationary phase reorganisation of the *Escherichia coli* transcription machinery by Crl protein, a fine-tuner of sigmas activity and levels. *EMBO J* **26**:1569–1578.
5. **Dong T, Kirchhof MG, Schellhorn HE.** 2008. RpoS regulation of gene expression during exponential growth of *Escherichia coli* K12. *Mol Genet Genomics* **279**:267–277.
6. **Dong T, Schellhorn HE.** 2009. Control of RpoS in global gene expression of *Escherichia coli* in minimal media. *Molecular genetics and genomics* : MGG **281**:19–33.
7. **Hengge-Aronis R.** 2002. Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. *Microbiology and molecular biology reviews* : MMBR **66**:373–95, table of contents.
8. **Klauck E, Typas A, Hengge R.** 2007. The sigmaS subunit of RNA polymerase as a signal integrator and network master regulator in the general stress response in *Escherichia coli*. *Science progress* **90**:103–27.
9. **Zhou Y, Gottesman S, Hoskins JR, Maurizi MR, Wickner S.** 2001. The RssB response regulator directly targets sigma(S) for degradation by ClpXP. *Genes & development* **15**:627–37.
10. **Jishage M, Ishihama A.** 1998. A stationary phase protein in *Escherichia coli* with binding activity to the major sigma subunit of RNA polymerase. *Proc Natl Acad Sci U S A* **95**:4953–4958.
11. **Mitchell JE, Oshima T, Piper SE, Webster CL, Westblade LF, Karimova G, Ladant D, Kolb A, Hobman JL, Busby SJW, Lee DJ.** 2007. The *Escherichia coli* regulator of sigma 70 protein, Rsd, can up-regulate some stress-dependent promoters by sequestering sigma 70. *Journal of bacteriology* **189**:3489–95.

12. **Kim EY, Shin M-S, Rhee JH, Choy HE.** 2004. Factors influencing preferential utilization of RNA polymerase containing sigma-38 in stationary-phase gene expression in *Escherichia coli*. *Journal of microbiology* (Seoul, Korea) **42**:103–10.
13. **Wassarman KM.** 2007. 6S RNA: a regulator of transcription. *Molecular microbiology* **65**:1425–31.
14. **Arnqvist A, Olsén A, Pfeifer J, Russell DG, Normark S.** 1992. The Crl protein activates cryptic genes for curli formation and fibronectin binding in *Escherichia coli* HB101. *Mol Microbiol* **6**:2443–2452.
15. **Pratt LA, Silhavy TJ.** 1998. Crl stimulates RpoS activity during stationary phase. *Molecular Microbiology* **29**:1225–1236.
16. **Bougdour A, Lelong C, Geiselmann J.** 2004. Crl, a low temperature-induced protein in *Escherichia coli* that binds directly to the stationary phase sigma subunit of RNA polymerase. *J Biol Chem* **279**:19540–19550.
17. **Robbe-Saule V, Jaumouillé V, Prévost M-C, Guadagnini S, Talhouarne C, Mathout H, Kolb A, Norel F.** 2006. Crl activates transcription initiation of RpoS-regulated genes involved in the multicellular behavior of *Salmonella enterica* serovar *Typhimurium*. *Journal of bacteriology* **188**:3983–3994.
18. **Gaal T, Mandel MJ, Silhavy TJ, Gourse RL.** 2006. Crl facilitates RNA polymerase holoenzyme formation. *J Bacteriol* **188**:7966–7970.
19. **Robbe-Saule V, Lopes MD, Kolb A, Norel F.** 2007. Physiological effects of Crl in *Salmonella* are modulated by sigmaS level and promoter specificity. *Journal of bacteriology* **189**:2976–87.
20. **England P, Westblade LF, Karimova G, Robbe-Saule V, Norel F, Kolb A.** 2008. Binding of the unorthodox transcription activator, Crl, to the components of the transcription machinery. *The Journal of biological chemistry* **283**:33455–64.
21. **Robbe-Saule V, Carreira I, Kolb A, Norel F.** 2008. Effect of growth temperature on Crl-dependent regulation of sigmaS activity in *Salmonella enterica* serovar *Typhimurium*. *Journal of bacteriology* **190**:4453–4459.
22. **Typas A, Becker G, Hengge R.** 2007. The molecular basis of selective promoter activation by the sigmaS subunit of RNA polymerase. *Molecular microbiology* **63**:1296–306.
23. **Lelong C, Aguiluz K, Luche S, Kuhn L, Garin J, Rabilloud T, Geiselmann J.** 2007. The Crl-RpoS regulon of *Escherichia coli*. *Mol Cell Proteomics* **6**:648–659.
24. **Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H.** 2006. Construction of *Escherichia coli* K-12 in-

- frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**:2006.0008.
25. **de Jong H, Ranquet C, Ropers D, Pinel C, Geiselmann J.** 2010. Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria. *BMC Syst Biol* **4**:55.
 26. **Boyer F, Besson B, Baptist G, Izard J, Pinel C, Ropers D, Geiselmann J, de Jong H.** 2010. WellReader: a MATLAB program for the analysis of fluorescence and luminescence reporter gene data. *Bioinformatics* **26**:1262–1263.
 27. **Lacour S, Kolb A, Boris Zehnder AJ, Landini P.** 2002. Mechanism of specific recognition of the *aidB* promoter by sigma(S)-RNA polymerase. *Biochemical and biophysical research communications* **292**:922–30.
 28. **Lacour S, Landini P.** 2004. SigmaS-dependent gene expression at the onset of stationary phase in *Escherichia coli*: function of sigmaS-dependent genes and identification of their promoter sequences. *J Bacteriol* **186**:7186–7195.
 29. **Azam TA, Iwata A, Nishimura A, Ueda S, Ishihama A.** 1999. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J Bacteriol* **181**:6361–6370.
 30. **Colland F, Barth M, Hengge-Aronis R, Kolb A.** 2000. sigma factor selectivity of *Escherichia coli* RNA polymerase: role for CRP, IHF and Lrp transcription factors. *The EMBO journal* **19**:3028–37.
 31. **Ihssen J, Egli T.** 2004. Specific growth rate and not cell density controls the general stress response in *Escherichia coli*. *Microbiology* **150**:1637–1648.
 32. **Hengge-Aronis R.** 1999. Interplay of global regulators and cell physiology in the general stress response of *Escherichia coli*. *Current opinion in microbiology* **2**:148–52.
 33. **Bokal AJ, Ross W, Gourse RL, IV AB.** 1995. The transcriptional activator protein FIS: DNA interactions and cooperative interactions with RNA polymerase at the *Escherichia coli* *rrnB* P1 promoter. *Journal of Molecular Biology* **245**:197–207.
 34. **Muskhelishvili G, Travers a a, Heumann H, Kahmann R.** 1995. FIS and RNA polymerase holoenzyme form a specific nucleoprotein complex at a stable RNA promoter. *The EMBO journal* **14**:1446–52.
 35. **Opel ML, Aeling KA, Holmes WM, Johnson RC, Benham CJ, Hatfield GW.** 2004. Activation of transcription initiation from a stable RNA promoter by a Fis protein-mediated DNA structural transmission mechanism. *Mol Microbiol* **53**:665–674.

36. **Bradley MD, Beach MB, de Koning APJ, Pratt TS, Osuna R.** 2007. Effects of Fis on *Escherichia coli* gene expression during different growth stages. *Microbiology* **153**:2922–2940.
37. **McLeod SM, Xu J, Cramton SE, Gaal T, Gourse RL, Johnson RC.** 1999. Localization of amino acids required for Fis to function as a class II transcriptional activator at the RpoS-dependent proP P2 promoter. *Journal of molecular biology* **294**:333–46.
38. **Typas A, Stella S, Johnson RC, Hengge R.** 2007. The -35 sequence location and the Fis-sigma factor interface determine sigmas selectivity of the proP (P2) promoter in *Escherichia coli*. *Molecular microbiology* **63**:780–96.
39. **Lange R, Hengge-Aronis R.** 1994. The cellular concentration of the sigma S subunit of RNA polymerase in *Escherichia coli* is controlled at the levels of transcription, translation, and protein stability. *Genes & Development* **8**:1600–1612.
40. **Hengge R.** 2008. The two-component network and the general stress sigma factor RpoS (sigma S) in *Escherichia coli*. *Advances in experimental medicine and biology* **631**:40–53.
41. **Ball CA, Osuna R, Ferguson KC, Johnson RC.** 1992. Dramatic changes in Fis levels upon nutrient upshift in *Escherichia coli*. *Journal of bacteriology* **174**:8043–56.

LEGENDS TO FIGURES

Figure 1. Dynamical expression pattern of the *crl* gene during growth.

The solid lines and the dashed lines respectively represent the luciferase reporter concentration and the absorbance in the different conditions (A) *crl* expression is transient after the dilution of an overnight culture into fresh minimal medium. The induction always appears at the same time during exponential phase independently of the time needed to enter stationary phase controlled by the glucose concentration. (B) The maximal transcriptional activity of *crl* always occurs for an absorbance of the culture between 0.2 and 0.3 independently of the inoculum dilution.

Figure 2. The *crl* promoter is σ^{70} -dependent

(A) Gel retardation assays performed with reconstituted RNA polymerases containing either σ^S or σ^{70} . With the *crl* promoter, retarded species were observed only with the σ^{70} holoenzyme even at the lowest concentration (17 nM). The σ^S -dependent *aidB* promoter was used as a reference. (B) Permanganate footprint on the non coding strand of *Pcrl*. (C) The sensitive T residues of the transcription bubble formed by E σ^{70} extended from positions -11 to -6 according to the TSS.

Figure 3. Activation of *crl* transcription by the nucleoprotein FIS.

The reporter concentration of the *crl::lux* fusion (solid lines), is lower in the *fis* mutant strain (green) as compared to the wild-type strain (blue). The transitory induction peak detected after 200 min of growth in the wild-type strain almost disappears in the *fis* mutant strain. Dashed lines represent the absorbance of each culture.

Figure 4. CRP represses *crl* transcription.

The solid lines represent the reporter concentration per cell and the absorbance at 600nm of the cultures are represented by the dashed lines. Transcription of *crl* is higher in both the *crp* (red) and the *cyaA* (orange) mutant strains as compared to the wild-type strain (blue).

Figure 5. The Fis protein binds to the *crl* promoter but the CRP protein does not.

These gel retardation assays were performed with purified protein in 200mM K-glutamate buffer, in the presence of cAMP (200 μ M) to test CRP binding. (A) Fis

binds *Pcrl* with a good affinity even at low temperature. (B) The binding of CRP was assayed using *PosmY* as a positive control (*osmY* is repressed by CRP). These assays were also performed in KCl (100 mM) and give very similar results.

Figure 6. The protein FIS interacts with nucleotide upstream of the -35 promoter element. The DNase I footprint was assayed on each strand of the *crl* promoter. (A) Both the template and non-template strands were incubated with increasing concentrations of purified Fis protein and treated with diluted DNaseI. The asterisks (*) point to hyper-reactive bands while the protected sites are marked with a “p”. (B) Sequence of the *crl* promoter localization of the reactive or protected nucleotides. The -10 and -35 promoter elements as well as the TSS are indicated in capital letters and serves as reference for nucleotide numbering. Bolded characters indicated nucleotide sensitive to Fis binding: underlined nucleotides are protected by Fis binding and the asterisks indicate hyper-reactivity induced by the Fis protein.

Figure 1

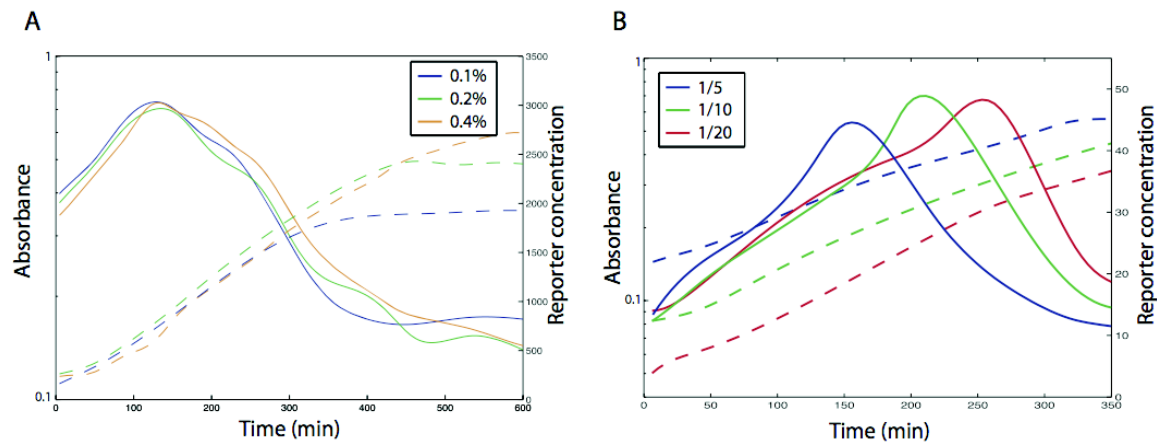


Figure 2

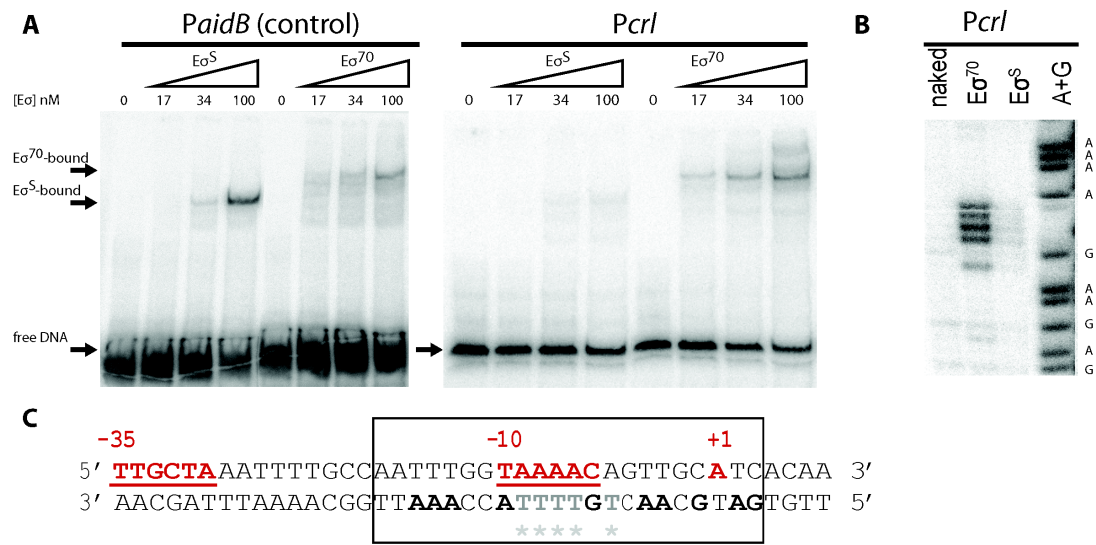


Figure 3

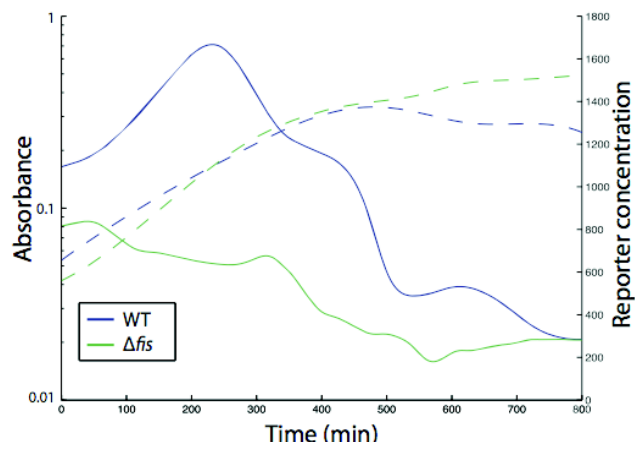


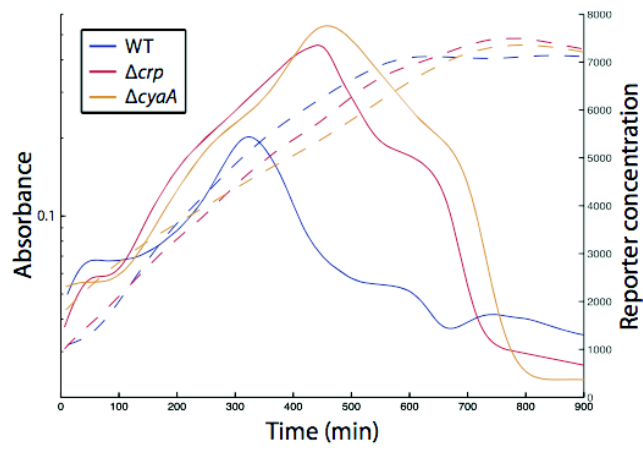
Figure 4

Figure 5

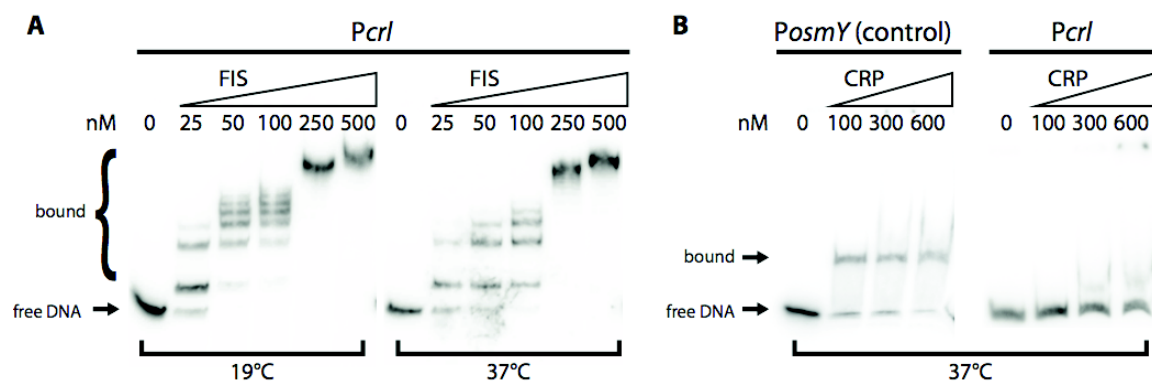
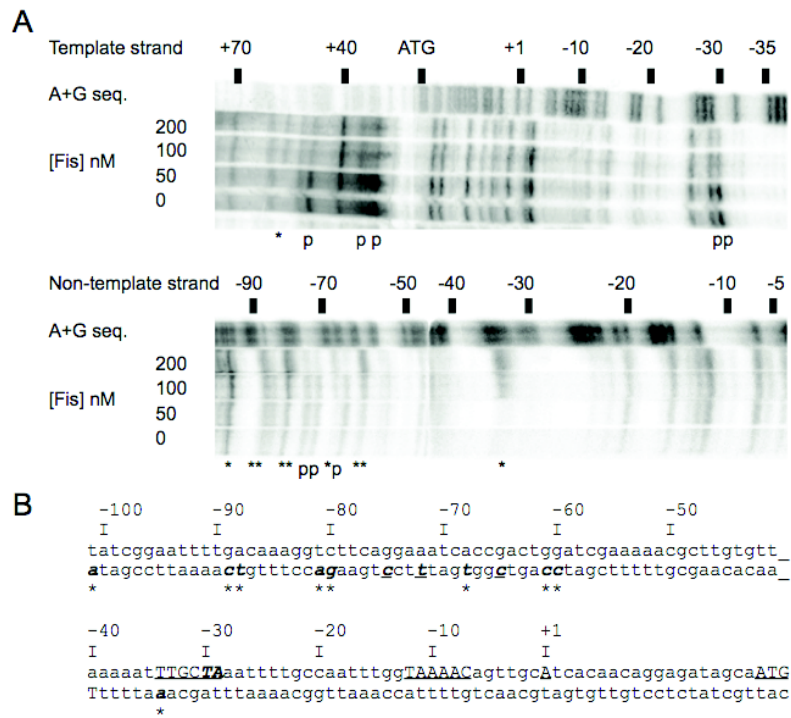


Figure 6

SUPPORTING INFORMATION

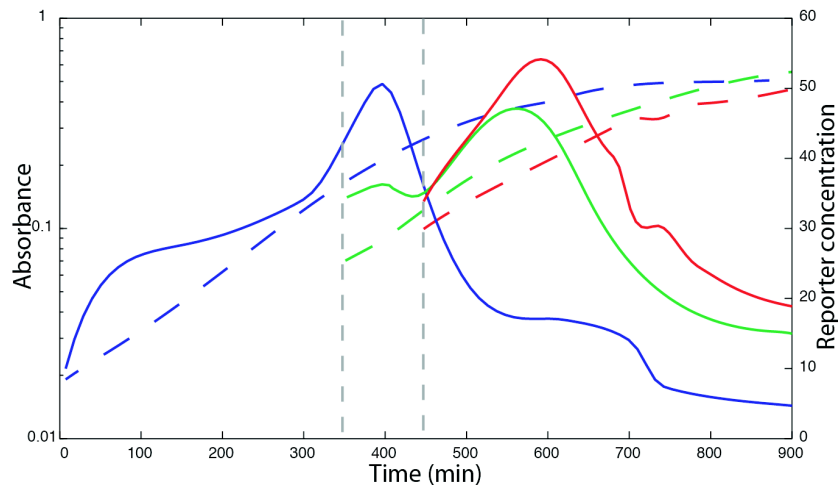


Figure S1. Induction peak of *crl* gene after a dilution in fresh medium.

The solid lines and the dashed lines respectively represent the luciferase reporter concentration and the absorbance in the different conditions. As in Figure 1 A, the *crl* expression is transient after the dilution of an overnight culture into fresh minimal medium (blue lines). After 350min, cells are diluted 4 times into fresh medium which results in a novel induction of *crl* expression (in green). Hundred minutes later, a third 2 times dilution is performed from the previous diluted culture (red). Each new dilution results in a novel transcriptional induction of *crl* independently from the growth phase: from stationary phase inoculum (blue) or from cells in active growth (green and red). The maximal transcriptional activity of the *crl* gene always occurs for an absorbance of the culture between 0.2 and 0.3.

Table S1. List of knockout mutants of the Keio collection used to measure *crl* transcription

<i>Gene name</i>	<i>Gene function</i>
A - Two-component systems	
<i>arcA/arcB</i>	Response to changing respiratory conditions of growth and regulator of gene expression under microaerobic and anaerobic conditions.
<i>atoS/atoC</i>	Induce the <i>ato</i> operon (<i>atoDAEB</i>) for metabolism of short chain fatty acids in response to the presence of acetoacetate.
<i>baeR/baeS</i>	Induce the expression of <i>spy</i> in response to envelope stresses such as spheroblast formation or misfolded pilus subunits.
<i>barA/uvrY</i>	Believed to play a role in central carbon metabolism via its regulation of the small non-coding RNAs <i>CsrC</i> and <i>CsrB</i> .
<i>basR/basS</i>	Response to excess external iron.
<i>cheA/cheY/cheB</i>	Chemotaxis
<i>citA/citB</i>	Activates genes involved in anaerobic citrate catabolism
<i>creB/creC</i>	Regulates the expression of various genes when cells are fermenting glycolytic carbon sources in minimal media and during aerobic growth on gluconeogenic carbon sources.
<i>cusR/cusS</i>	Respond to elevated copper and silver concentrations.
<i>cpxA/cpxR</i>	Respond to cell envelope damage.
<i>dcuR/dcuS</i>	Controls the expression of genes relating to the catabolism of exogenous C4 dicarboxylates (aspartate, fumarate, malate and succinate).
<i>envZ/ompR</i>	Respond to changes in medium osmolarity.
<i>evgA/evgS</i>	Believed to initiate a complex transcriptional cascade that functions for acid and drug resistance.
<i>glnG/glnL</i>	Detect nitrogen availability.
<i>kdpD/kdpE</i>	Detect change in the intracellular K concentration and ionic strength.
<i>narX/narL</i>	Control anaerobic respiratory gene expression in response to nitrate and nitrite.
<i>narQ/narP</i>	Control anaerobic respiratory gene expression in response to nitrate and nitrite.
<i>phoP/phoQ</i>	Respond to low levels of extracellular Mg^{2+} .
<i>phoR/phoB</i>	Respond to variations in the level of extracellular inorganic phosphate (Pi).
<i>qseB/qseC</i>	Thought to be a quorum sensing two-component system involved in regulation of flagella biogenesis.
<i>rstB/rstA</i>	Thought to be involved in the adaptive response to acid conditions
<i>torS/torR</i>	Respond to pH stress.
<i>uhpA/uhpB</i>	Responsible for sensing and transport of the metabolic intermediate, glucose-6-phosphate.
<i>yedV/yedW</i>	Putative two-component system.
<i>yehT/yehU</i>	Putative two-component system.
<i>yfhK/yfhA</i>	Putative two-component system.

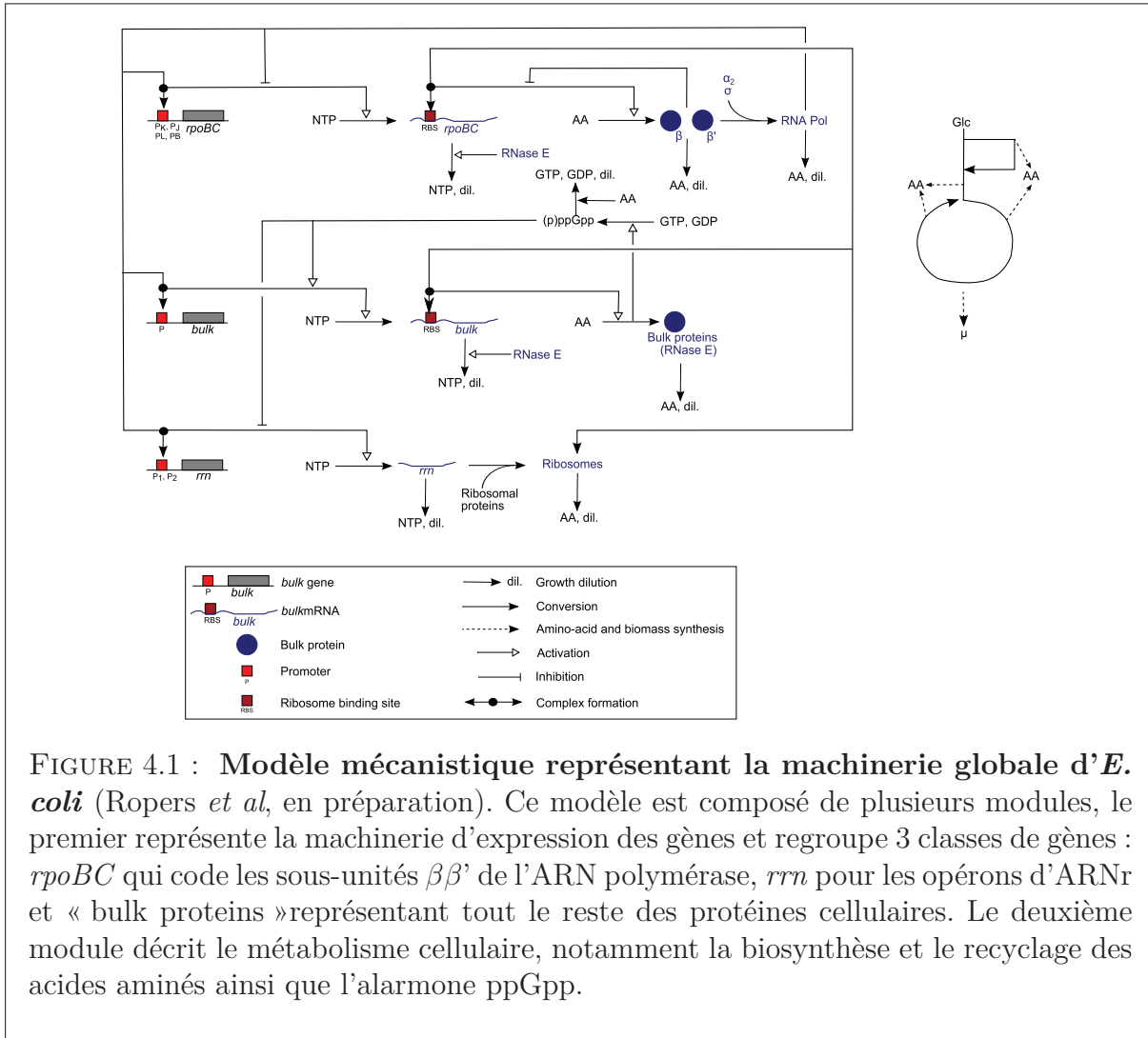
3. ÉTUDE DE L'EXPRESSION DE RÉGULATEURS GLOBAUX : CAS DU GÈNE *crl*

<i>ypdA/ypdB</i>	Putative two-component system.
<i>zraR/zraS</i>	Respond to high external zinc concentration.
<i>rcaA/rcaB</i>	Respond to osmotic shock, desiccation, and cell membrane integrity perturbations.
B – Genes involved in Quorum Sensing	
<i>lsrB/lsrD/lsrG</i>	Predicted ABC transporter to transport the quorum sensing autoinducer molecule AI-2.
<i>luxS</i>	Synthase that catalyzes formation of AI-2.
<i>sdiA</i>	Homologous to quorum-sensing transcription factors that belong to the LuxR family.
<i>tnaA</i>	Tryptophanase which catalyses the hydrolytic cleavage of L-tryptophan to indole.
<i>wrbA</i>	Tryptophan (W) repressor-binding protein
<i>ygiU</i>	Links the AI-2 quorum sensing signal to biofilm formation
<i>pfs</i>	Required for AI-2 production in <i>Salmonella enterica</i> .
<i>trpC/trpR</i>	Enzymes implied in the tryptophan biosynthesis pathway.
C – Global regulators and nucleoproteins	
<i>crp</i>	Global regulator working with cAMP.
<i>cyaA</i>	Adenylate cyclase, enzyme which produces the cAMP
<i>cbpA</i>	Co-chaperone with DnaK
<i>dps</i>	DNA-binding protein
<i>fis</i>	DNA-binding and bending protein. Also a transcriptional regulator (regulates rRNA and tRNA)
<i>fnr</i>	Transcriptional regulator that mediates the transition from aerobic to anaerobic growth.
<i>hfq</i>	RNA-binding protein and RNA chaperone.
<i>hns</i>	Histone-like nucleoid structuring protein
<i>hupA/hupB</i>	Code for HU protein, a histone-like nucleoid structuring protein
<i>ihfA/ihfB</i>	Code for IHF subunits. IHF is a global regulatory protein. It binds DNA at specific site and helps to maintain DNA architecture.
<i>Lrp</i>	Leucine-responsive regulatory protein, a global regulator.
<i>stpA</i>	"Suppressor of td phenotype A," a nucleoid-associated multifunctional protein. Paralogous of HN-S

Effet de la machinerie globale d'expression sur la croissance

4.1 Rôle de la machinerie globale sur l'expression génique

Dans le chapitre précédent, nous avons étudié l'expression du régulateur *crl*. Nous avons montré que son expression est transitoire et présente un pic d'activité en phase exponentielle, lorsque des cellules provenant d'une phase stationnaire sont placées dans du milieu frais. Les régulateurs CRP et la nucléoprotéine Fis régulent son expression, mais ces régulations spécifiques ne suffisent pas à expliquer entièrement le profil de *crl*. En effet, en l'absence de ces régulateurs on observe toujours un pic d'expression de *crl* de moindre amplitude. Au laboratoire, d'autres études ont montré que l'expression de plusieurs régulateurs globaux, comme RpoS, les topoisomérases GyrA, GyrB et TopA, CRP ou encore la nucléoprotéine Fis présentent un pic d'activité lors de la phase exponentielle. Leur profils d'expression sont très similaires à celui de *crl*. Ces résultats suggèrent un mécanisme de régulation commun à tous ces régulateurs. Nous avons alors émis l'hypothèse que cette expression transitoire pourrait être induite par la machinerie globale d'expression, c'est-à-dire par l'ARN polymérase et les ribosomes de la cellule. En effet, ces composants majeurs de la cellule présentent des concentrations et niveaux d'activité différents lorsque le taux de croissance change (voir Tableau 1.1 et Fig. 2.1).



Afin de vérifier cette hypothèse, un modèle cinétique simplifié de la machinerie d'expression génique a été développé au laboratoire (Fig. 4.1 ; Ropers *et al*. en préparation). Dans ce modèle, les composants de la cellule sont regroupés en trois grandes classes : l'ARN polymérase (gènes *rpoB* et *rpoC* codant les sous-unités $\beta\beta'$ dont l'expression est limitante pour la formation d'ARN polymérase), les ribosomes (opérons *rrn* codant pour les ARN stables, dont le taux de synthèse détermine la vitesse d'accumulation des protéines ribosomales et donc, des ribosomes) et les protéines cellulaires (gènes « bulk »). Ces composants interagissent via différentes boucles de rétroaction : une rétroaction positive de l'ARN polymérase et des ribosomes sur tous les ARN et protéines, respectivement ; l'autorégulation de l'expression de l'ARN polymérase ; la répartition de l'ARN polymérase et des ribosomes entre leurs formes libres et liées ; le couplage transcription-traduction et

traduction-dégradation ; l'effet de petites molécules (ppGpp) et des métabolites (acides aminés) sur la traduction. Ce modèle est capable de reproduire quantitativement les données d'expression d'un certain nombre de régulateurs globaux décrites précédemment.

Ainsi, des remaniements de la quantité et de l'activité des composants de la machinerie d'expression génique permettent d'expliquer la formation de pics d'expression protéique lors d'une reprise de croissance. Ces résultats sont cohérents avec les travaux de la littérature qui expliquent le lien étroit entre taux de croissance au steady-state et composition macromoléculaire de la cellule sur la base de changements d'activités transcriptionnelle et traductionnelle (cf. Section 1.3 de l'introduction, page 13).

4.2 Contrôle du taux de croissance en modulant la machinerie globale

Pourquoi, dans des conditions transitoires comme à l'équilibre, l'activité de la machinerie cellulaire s'adapte-t-elle au taux de croissance ? Est-ce que la réciproque est vraie ? Peut-on moduler le taux de croissance en agissant sur la machinerie d'expression génique ? Quelles seraient pour la bactérie les conséquences physiologiques et moléculaires d'un tel contrôle de la croissance ? Afin de répondre à ces questions, la seconde partie de ma thèse s'est focalisée sur le développement d'une souche d'*Escherichia coli* permettant de contrôler le taux de croissance en affectant la machinerie d'expression génique. Deux systèmes de contrôle différents ont été envisagés : le premier cible l'activité traductionnelle de la bactérie en inactivant les ribosomes, tandis que le second cible l'activité transcriptionnelle en ajustant la quantité intracellulaire d'ARN polymérase.

4.2.1 Modulation de l'activité traductionnelle

Notre première approche pour contrôler le taux de croissance d'*E. coli* consistait à modifier son activité traductionnelle. En effet la croissance nécessite la production d'une grande quantité de protéines. On peut alors penser que si l'on affecte la quantité de ribosomes, cela aura des répercussions sur le taux de croissance. Nous avons vu dans l'introduction que les ribosomes sont inactivés par les protéines RMF et HPF. Les ribosomes se

trouvent alors sous forme de dimères formant des particules 100S et n'ont plus aucune activité traductionnelle. Seule une partie des ribosomes reste donc sous leur forme active afin de permettre l'expression de protéines nécessaires à la survie en condition de stress nutritionnel (Fig. 2.7). Nous avons donc testé l'effet des mutations de ces 3 facteurs, ainsi que la surexpression de RMF sur le niveau de traduction de la cellule et sur sa croissance.

Afin de mesurer l'activité traductionnelle de la cellule, nous avons construit un plasmide - appelé pZa-lux - exprimant la luciférase. Cette expression est sous le contrôle d'un promoteur inductible par l'anhydrotetracycline (atc). Lorsque l'on ajoute de l'atc dans le milieu, le répresseur TetR s'y fixe et ne réprime plus le promoteur permettant ainsi l'expression de la luciférase et donc, la production de lumière. L'expression de la luciférase est alors directement proportionnelle à la concentration d'atc ajoutée dans le milieu. On peut dès lors utiliser ce plasmide comme un rapporteur de l'activité traductionnelle de la cellule. Pour une même concentration d'atc, peu de luciférase sera exprimée si la traduction est peu efficace et le niveau de lumière mesuré sera faible. A contrario, si la traduction est très active, le signal mesuré sera plus fort.

Nous avons donc transformé et mesuré l'expression de ce plasmide dans les différents mutants défailants pour la formation des dimères 100S, provenant de la collection Keio (Baba *et al.*, 2006). Nous avons ensuite comparé les profils obtenus avec celui de la souche sauvage (Fig. 4.2 A). Il apparaît que le mutant *hpf*, dont le produit permet la formation de particules 90S, possède une plus faible luminescence que la souche sauvage. Ceci est contradictoire avec la fonction de HPF qui est d'inactiver les ribosomes. On s'attendrait, en effet, à avoir une plus forte activité traductionnelle sans cette protéine et donc une luminescence plus forte. Concernant le gène *rmf*, qui lui aussi permet la dimérisation des ribosomes, on observe une légère augmentation de luminescence comparativement à la souche sauvage. Ces deux résultats opposés pour des protéines ayant des rôles proches n'étaient pas très encourageants. De plus les mutations n'ont aucun effet sur le taux de croissance même si elles sont décrites comme diminuant la survie en phase stationnaire (Yamagishi *et al.*, 1993). Le mutant *yfiA*, dont la protéine se fixe sur les ribosomes 70S et favorise ainsi leur forme active, semble affecter de façon modéré l'expression de la luciférase. Son expression est à un niveau intermédiaire entre la souche sauvage et le mutant *hpf*. Ce résultat semble incohérent avec la fonction de YfiA et, malgré cette baisse

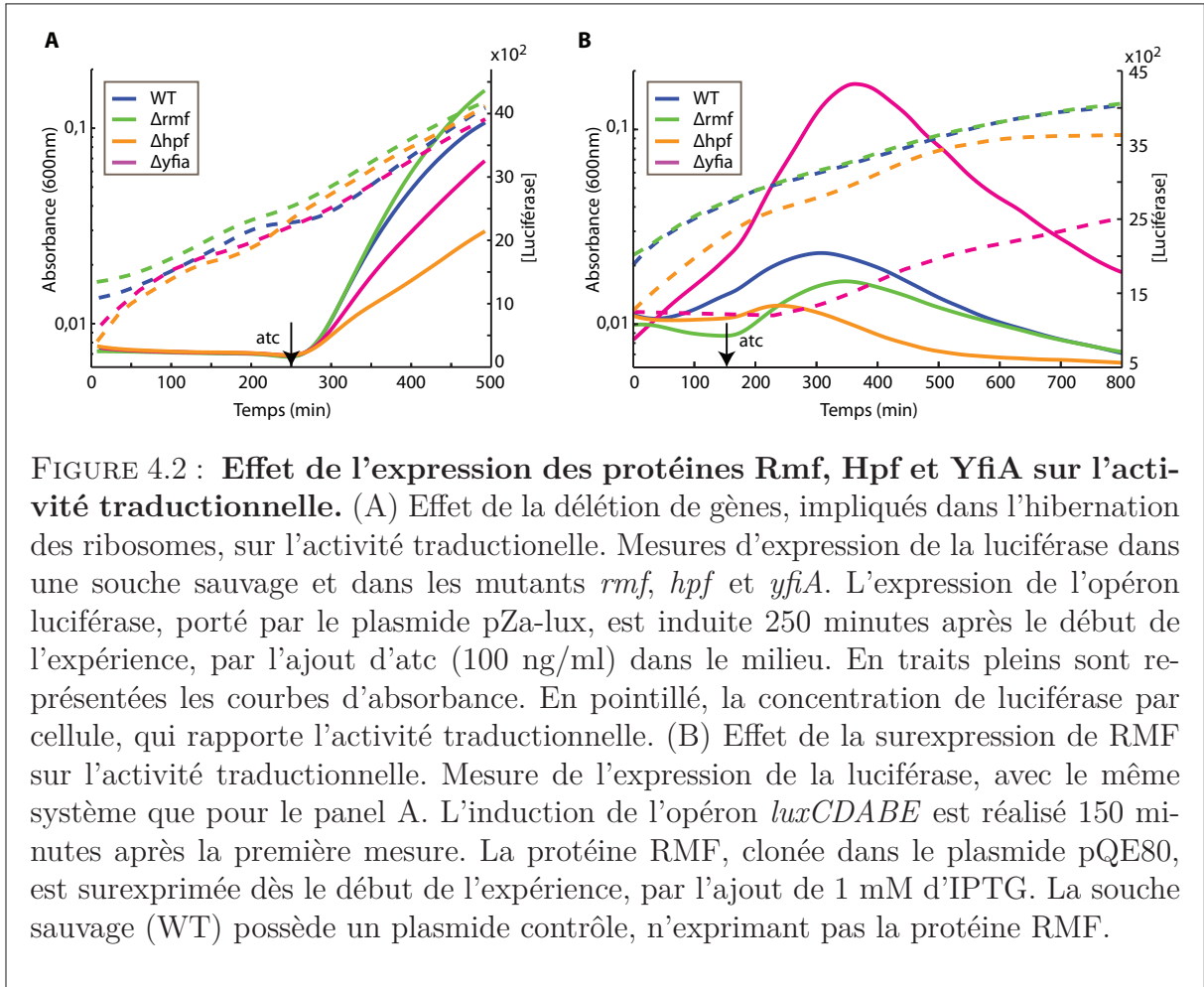


FIGURE 4.2 : Effet de l'expression des protéines Rmf, Hpf et YfiA sur l'activité traductionnelle. (A) Effet de la délétion de gènes, impliqués dans l'hibernation des ribosomes, sur l'activité traductionnelle. Mesures d'expression de la luciférase dans une souche sauvage et dans les mutants *rmf*, *hpf* et *yfiA*. L'expression de l'opéron luciférase, porté par le plasmide pZa-lux, est induite 250 minutes après le début de l'expérience, par l'ajout d'atc (100 ng/ml) dans le milieu. En traits pleins sont représentées les courbes d'absorbance. En pointillé, la concentration de luciférase par cellule, qui rapporte l'activité traductionnelle. (B) Effet de la surexpression de RMF sur l'activité traductionnelle. Mesure de l'expression de la luciférase, avec le même système que pour le panel A. L'induction de l'opéron *luxCDABE* est réalisé 150 minutes après la première mesure. La protéine RMF, clonée dans le plasmide pQE80, est surexprimée dès le début de l'expérience, par l'ajout de 1 mM d'IPTG. La souche sauvage (WT) possède un plasmide contrôle, n'exprimant pas la protéine RMF.

d'activité, la croissance n'est quant à elle pas affectée. De plus, ces protéines s'accumulent préférentiellement lors de la phase stationnaire. Il est donc fort probable que les mutations simples ne suffisent pas à affecter la traduction au point de diminuer le taux de croissance.

Les mutations ne montrant pas d'effet sur la croissance, nous avons donc cloné la protéine RMF dans un plasmide de surexpression : pQE80 (commercialisé par Qiagen), afin d'induire son expression en ajoutant de l'IPTG dans le milieu. Avec ce système il est donc possible de produire RMF lors de la phase exponentielle, alors qu'elle n'est naturellement présente qu'en phase stationnaire. Nous avons donc exprimé cette protéine dans les mutants *rmf*, *hpf* et *yfiA*. En parallèle, nous avons utilisé un plasmide contrôle, ne possédant pas *rmf*, dans la souche sauvage. Ces souches ont été toutes transformées par le plasmide pZa-lux, dont nous avons mesuré l'activité traductionnelle (Fig. 4.2 B). L'expression de RMF a été induite dès le début de l'expérience en ajoutant de l'IPTG dans

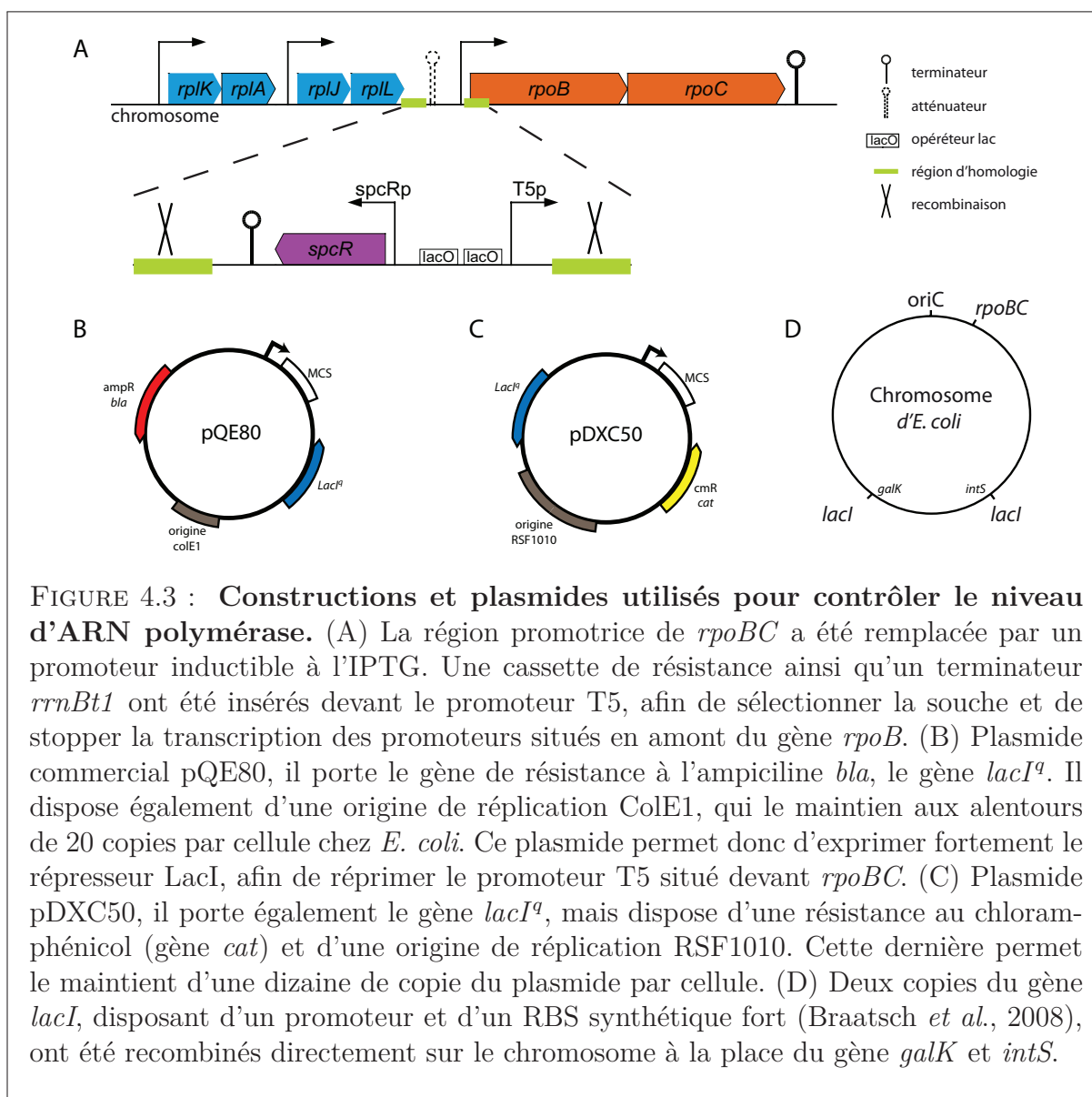
le milieu. Tandis que celle de l'opéron luciférase, 150 minutes plus tard, lors de la phase exponentielle en ajoutant de l'atc. Cette surexpression de RMF provoque une baisse de la luminescence chez le mutant *hpf*, ce qui suggère une baisse de la traduction de l'opéron luciférase. Cette diminution serait cohérente avec la formation de dimère de ribosomes 90S et donc une inactivation partielle des ribosomes lors de la phase exponentielle. Malgré cette potentielle baisse d'activité traductionnelle, la croissance n'est pas affectée. Dans le mutant *rmf*, la surexpression de RMF semble restaurer l'expression de pZa-lux observée dans la souche sauvage. La surexpression de RMF dans le mutant *yfiA* montre un résultat inattendu. En effet l'activité de pZa-lux y est deux fois plus forte que dans la souche sauvage et la croissance de cette souche est très nettement diminuée. Ce résultat semble incohérent avec la fonction d'inactivation des ribosomes de RMF.

L'ensemble de ces résultats ont permis de voir qu'il était difficile de contrôler la croissance en modulant l'inactivation des ribosomes. En effet, si leur fonctions ont été identifiées, la régulation de leur expression est encore mal connue. Il est vraisemblable que la mutation ou la surexpression de ces protéines soit compensée par des mécanismes de régulation encore inconnus visant à maintenir une certaine proportion de ribosomes actifs et inactifs dans les cellules."

4.2.2 Modulation de l'activité de l'ARN polymérase

N'ayant pu contrôler la croissance en modulant la traduction, nous avons décidé d'utiliser une autre approche consistant à moduler l'activité transcriptionnelle. En effet chez *Escherichia coli*, il existe une ARN polymérase unique composée de plusieurs sous-unités. Les sous-unités sigma étant nombreuses et donnant la spécificité à l'enzyme, nous les avons éliminées en tant que cibles. Le cœur de l'enzyme semblait un bien meilleur candidat. En effet, l'expression des sous unités β et β' est connue pour être limitante dans la production de l'enzyme cœur, et leurs gènes *rpoB* et *rpoC* sont en opéron (Fig. 2.2).

Nous avons donc choisi de mettre l'expression de cet opéron *rpoBC* sous le contrôle d'un promoteur IPTG inductible. La souche d'*E. coli* portant cette construction sur le chromosome a été appelée R pour « regulated » (Fig. 4.3). En absence d'IPTG, le promoteur est réprimé par la protéine LacI et l'expression des sous-unités $\beta\beta'$ est abolie. A l'inverse, lorsque l'on ajoute de l'IPTG dans le milieu, LacI vient se complexer avec ce



dernier et n'empêche plus la transcription de *rpoBC*. La quantité intracellulaire de LacI est critique pour le bon fonctionnement de notre système. Au cours de ma thèse, j'ai testé 2 stratégies pour maintenir la quantité intracellulaire de LacI.

Système de répression plasmidique

Dans un premier temps, il est apparu que le gène *lacI* natif d'*E. coli* ne suffisait pas pour réprimer l'expression de $\beta\beta'$. Nous avons donc décidé d'utiliser le plasmide pQE80, portant le gène *lacI^q* (Fig. 4.3 B). Ce gène est une version améliorée de *lacI*, il dispose d'une meilleure expression grâce à un codon initiateur de la traduction consensus. En

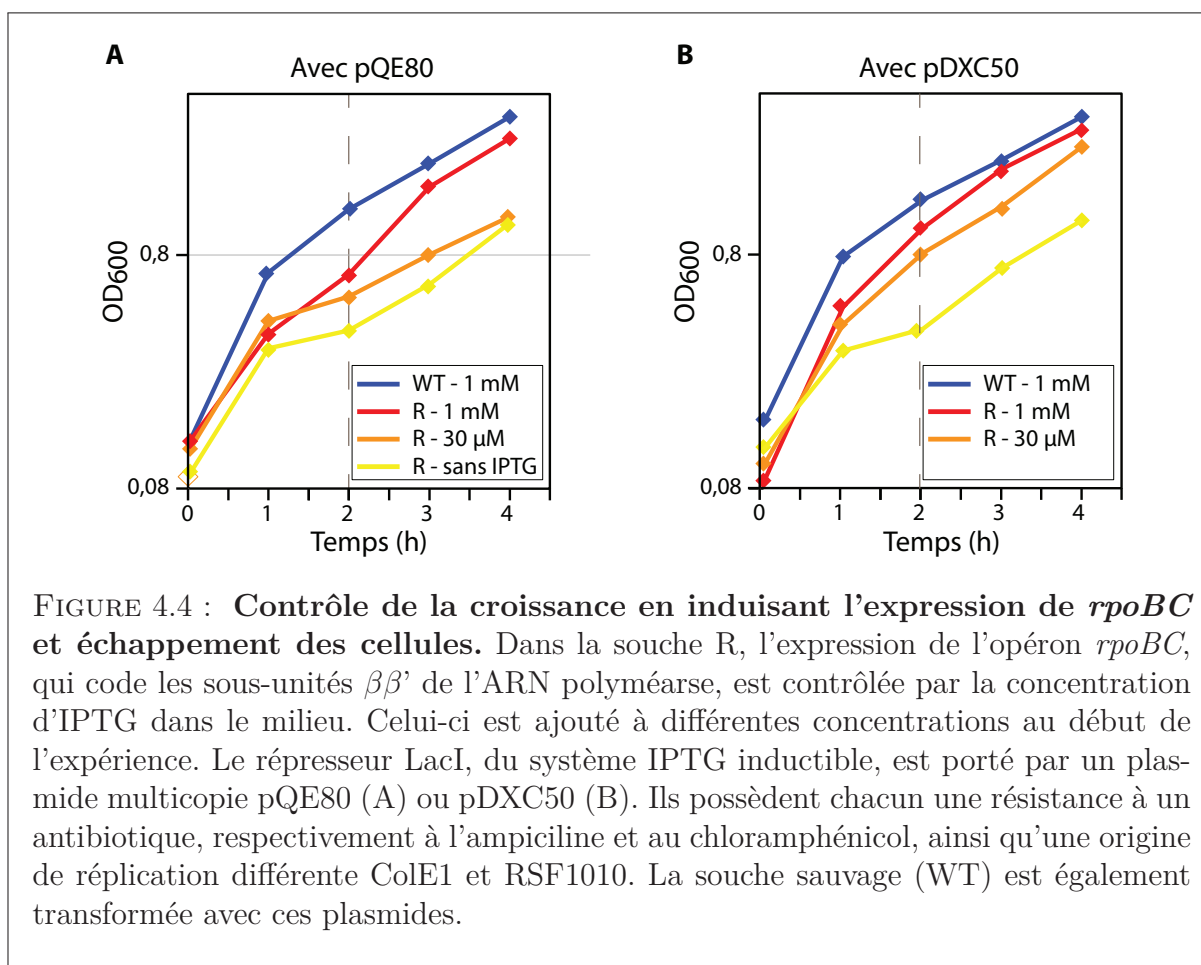
outre, le fait que le plasmide soit en plusieurs copies augmente son expression et diminue la possibilité de fixer une mutation inactivant le gène *lacI*.

Avec ce système nous avons montré, qu'avec une forte concentration d'IPTG, les bactéries modifiées présentent le même taux de croissance qu'une souche sauvage. A l'opposé, lorsqu'il n'y a pas d'IPTG dans le milieu, c'est-à-dire lorsque les sous unités $\beta\beta'$ ne sont plus détectables par western blot, les bactéries stoppent leur croissance. Nous avons ainsi montré que le taux de croissance des cellules répondait de façon dose dépendante à la concentration d'IPTG présente dans le milieu (Fig. 4.4).

Cependant, après deux heures de croissance sans IPTG, nous avons observé un échappement de la population comme si notre système perdait de son efficacité : en effet les cellules retrouvent une croissance proche de celle observée chez la souche sauvage (Fig. 4.4 A). Nous avons montré par séquençage que le promoteur situé en amont de *rpoBC* n'avait pas fixé de mutation. Cet échappement ne venant pas d'une mutation sur le promoteur, il était également possible qu'une mutation sur le gène *lacI* puisse être à l'origine de la dé-répression de notre système. Le fait qu'il soit en plusieurs copies n'allait cependant pas dans ce sens. Nous avons donc vérifié si le plasmide était toujours présent dans les cellules, une fois qu'elles avaient repris une croissance normale.

Il s'est avéré que les cellules ne possèdent plus le plasmide portant le répresseur *lacI*, malgré la présence d'ampiciline qui normalement permet son maintien dans la cellule. Cependant le système de résistance à l'ampiciline est connu pour perdre de son efficacité au cours du temps. En effet le gène *bla*, porté par le plasmide pQE80, code pour une beta-lactamase qui dégrade l'ampiciline présente dans le milieu. Il est donc raisonnable de penser qu'après un certain temps, la concentration d'ampiciline est tellement faible qu'il n'est plus nécessaire de porter la résistance à l'ampiciline pour croître. De plus, le système de réplication du plasmide pQE80 est dépendant de la transcription : il faut en effet un primer ARN pour initier sa réplication. Lorsque la souche produit peu d'ARN polymérase, en présence de faible quantité d'IPTG, elle est susceptible de moins transcrire cet ARN.

Ces deux mécanismes pourraient expliquer pourquoi le plasmide n'est plus maintenu lorsque les cellules sont forcées d'exprimer peu d'ARN polymérase. La pression de sélection sur notre système étant alors tellement forte, les cellules perdent le plasmide donc le



répresseur LacI. Elles retrouveraient alors une forte expression de l'ARN polymérase et un taux de croissance normal.

Pour palier à ce problème d'échappement, nous avons utilisé un autre plasmide pDXC50 (Fig. 4.3 C). Il porte également le gène *lacI*, confère la résistance au chloramphenicol et utilise une origine de répllication RSF1010. Cette origine de répllication possède sa propre primase et hélicase contrairement à l'origine ColE1 du plasmide pQE80 (Honda *et al.*, 1991). Avec ce nouveau plasmide, nous avons pu reproduire l'effet de notre contrôle de *rpoBC* sur la croissance, mais nous avons malheureusement observé le même phénomène qu'avec le plasmide pQE80. Après deux heures de croissance sans IPTG, le taux de croissance de la population redevient le même que dans la souche sauvage et les cellules se sont débarrassées du plasmide (Fig. 4.4 B). Une explication possible serait que le niveau de transcription soit tellement bas qu'il empêcherait une expression suffisante du système de répllication des plasmides. Ceci expliquerait qu'avec deux systèmes de répression

plasmidique différents, nous observons le même problème d'échappement.

Système de répression chromosomique

Afin d'éviter ce phénomène de perte du répresseur, nous avons donc décidé de cloner sur le chromosome deux copies du gène *lacI*, possédant un promoteur et un RBS synthétiques consensus (Fig. 4.3 C). Ce nouveau système s'est avéré fonctionnel ; nous avons pu confirmer l'effet dose dépendant de l'expression de *rpoBC* sur la croissance et avons pu poursuivre nos investigations. Ces observations montrent que notre contrôleur fonctionne et qu'il ne semble pas y avoir de mécanisme compensant les faibles concentrations d'ARN polymérase. Nous avons donc par la suite utilisé cette souche pour identifier les conséquences de cette faible expression sur la physiologie cellulaire et sur l'expression globale du génome. Ces travaux font l'objet d'un article soumis prochainement, dont je résume les principaux points ci-dessous.

Voyant que les bactéries arboraient une croissance nulle en absence d'IPTG, nous avons cherché à savoir quelles fonctions cellulaires étaient affectées. En procédant à des étalements ainsi qu'en observant les cellules sous microscope, il est apparu qu'elles ne se divisaient plus et présentaient une morphologie anormale : elles formaient toutes de long filaments (Fig. 1 D et 3 A de l'article). Pour des concentrations intermédiaires d'IPTG, alors qu'on observe un taux de croissance intermédiaire en population, nous avons montré que la population de bactérie n'est plus homogène. Elle est alors composée de cellules filamenteuses et de cellules ayant une morphologie identique à celle observée chez la souche sauvage (Fig. 3 A de l'article).

La filamentation chez les bactéries est un phénomène souvent observé lorsqu'elles sont soumises à un stress, par exemple, lorsqu'elle subissent des dommages à l'ADN et que la réponse SOS est activée (Hill *et al.*, 1997), ou encore lorsque des bactéries pathogènes échappent à la réponse immunitaire de leur hôte (Justice *et al.*, 2006). Néanmoins, même lorsque des cellules filamentent, elles s'avèrent souvent capables de ségréger leur nucléoïde et donc de répliquer leur chromosome. Nous avons testé s'il en était de même dans notre souche, lorsqu'elle n'exprime plus de sous-unité $\beta\beta'$. Ainsi, à l'aide d'un marqueur fluorescent colorant l'ADN, nous avons montré que la réplication et la ségrégation des chromosomes ne sont pas affectées dans les cellules filamenteuses (Fig. 3 B de l'article).

Lorsqu'elles n'expriment plus d'ARN polymérase, les cellules ne se divisent plus mais continuent de s'allonger et de répliquer leur chromosome. Nous avons par la suite voulu savoir si le stress, provoqué par la très faible concentration d'ARN polymérase, engendrait un état physiologique irréversible ou si les cellules pouvaient retrouver une morphologie normale lorsque le stress disparaît. Pour cela, j'ai été accueilli 3 semaines dans l'équipe de F. Taddei à Paris. Durant cette collaboration, et avec l'aide de Y. Yang, j'ai pu utiliser leur système de micro-fluidique permettant de suivre la croissance à long terme de cellules individuelles. Nous avons alors constaté que 90% des bactéries, présentant une morphologie filamenteuse en absence d'IPTG, restauraient une croissance et une morphologie normale lorsque l'on en rajoutait dans le milieu (Fig. 4 de l'article).

Après avoir étudié la physiologie de notre souche, nous nous sommes intéressés à l'impact du niveau d'ARN polymérase sur l'expression du génome. Par une technique de séquençage des ARN messagers, nous avons mesuré l'expression de l'ensemble des gènes d'*E. coli*, dans la souche sauvage ainsi que dans notre souche pour 3 niveaux d'ARN polymérase différents. Il s'est avéré que l'ensemble des classes fonctionnelles de gène est affecté par notre système, à l'exception des gènes qui codent les protéines ribosomales (Fig. 5 de l'article).

Publication n°3 :

Growth control of *Escherichia coli* by the global transcriptional machinery

Growth control of *Escherichia coli* by the global transcriptional machinery

Jérôme Izard * [†] *et al.*

*Institut Jean Roget, LAPM, UMR5163, Campus Santé, Université Joseph Fourier, Domaine de la Merci, 38700 La Tronche, France, and [†]INRIA Grenoble - Rhône-Alpes, 655 Av. de l'Europe, Montbonnot, 38334 St Ismier Cedex, France

Submitted to Proceedings of the National Academy of Sciences of the United States of America

The maximal growth rate of a bacterial population is determined by the quality and quantity of the nutrients in the culture medium. Given this constraint, the organism optimizes physiological parameters such as gene expression and enzymatic activities for optimal utilization of these nutrients. We have engineered an *E. coli* strain for which the growth rate can be set experimentally between zero and the maximal growth rate by putting the transcription of an essential component of the global gene expression machinery, RNA polymerase, under the control of an inducible promoter. Growth of these bacteria can be arrested in a reversible manner. After shutting off production of RNA polymerase, the bacteria continue DNA replication and maintain metabolic activity, but no longer divide. The resulting filamentation is completely reversed when RNA polymerase is provided again. The analysis of the transcriptome of bacteria at growth rates restricted by the concentration of RNA polymerase reveals that all functional classes of genes are almost equally affected by the reduction in transcriptional activity. The genes coding for ribosomal proteins are a possible exception. Our results suggest that decreased transcription is at least partially compensated by increased translation.

growth | RNA polymerase

A major challenge for living organisms is the regulation of cellular growth in a changing environment. Sudden changes of nutrient availability or environmental stress conditions require rapid adjustment of the growth rate. In bacteria, such as the model organism *Escherichia coli*, the control of cellular growth also has important applications in biotechnology (1–3). For both, fundamental and applied research, it is therefore important to understand growth-rate control and to be able to adjust cellular growth to a desired level in a reversible manner.

Early work in bacterial physiology has characterized the adjustment of cell composition as a function of growth rate (4–6). These experiments have shown that physiological state of the cell is determined by the growth rate and *not* nature of the specific nutrient: cells growing at the same rate on different growth media have the same content in DNA, RNA, and protein (7). Among the global physiological parameters that vary with growth rate we find gene copy number, the intracellular concentration of RNA polymerase, ribosomes, the size of metabolic pools, ... (4–6). Changes in transcription and translation rates then set the nucleic and protein composition of cells to new values. While much research has addressed the question of how specific signals induce specific cellular responses, the mechanisms connecting growth rate to global physiological parameters remain largely unknown. Previous experiments have measured the changes in important physiological parameters, such as the concentration of RNA polymerase (4), as a function of an externally imposed growth rate or after genetic and metabolic perturbations that specifically inactivate cellular components involved in the biomass formation (8, 9).

In this manuscript, we take the inverse approach: we modify the concentration of RNA polymerase and observe the effects of this perturbation on growth rate and cell physiology. We put the transcription of the *rpoBC* operon, coding for the two large subunits ($\beta\beta'$) of RNA polymerase under the control of an inducible promoter. Addition of different concentration of the inducer, IPTG, to the growth medium

leads to a controlled growth rate independent of the quality of the carbon source. In the absence of inducer, the bacteria stop to grow.

This growth arrest is reversible, since the intracellular concentration of the $\beta\beta'$ subunits, as well as growth rate, increase with increasing concentrations of IPTG. Surprisingly, growth-arrested cells remain physiologically active: they are active in chromosome replication and they possess metabolic activity. Analysis of the global transcriptional profile by RNAseq suggests that the decrease in RNA polymerase concentration is compensated by increased translational activity. Despite the complexity of the network controlling bacterial growth in *E. coli*, a local and specific control of one key component of the global gene expression machinery is sufficient to set growth rate to a desired level and other cell parameters (metabolic activity, replication,...) are adjusted accordingly.

Results

Engineering RNA polymerase expression. The core RNA polymerase (RNA polymerase) from *Escherichia coli* consists of three subunits and has the composition $\alpha_2\beta\beta'$. The α subunit is produced in excess during growth, the amount of the $\beta\beta'$ subunits is limiting (10–12). These two subunits are the product of the *rpoB* and *rpoC* genes, organized in an operon with upstream genes coding for ribosomal proteins (*rplKALJL*) (13). The expression of the two genes is regulated both at transcription and translation by complex, and still incompletely understood mechanisms. An attenuator of transcription located in the intergenic region *rpoL-rpoB* decouples the expression of ribosomal proteins and the two RNA polymerase subunits (14, 15).

Since *Escherichia coli* only possesses one RNA polymerase, the production of this enzyme is essential for cell survival. When the activity of RNA polymerase is inhibited, for example by the action of an antibiotic such as rifampicin, the bacteria stop growing or even die (16). By adjusting the activity (or concentration) of RNA polymerase, we should therefore be able to vary growth continuously between zero and the maximal growth rate supported by a given carbon source. This, and the fact that the two $\beta\beta'$ subunits are limiting for the production of the core enzyme (17), suggests that the bacterial growth rate could be adjusted by controlling the transcription of *rpoBC*.

We therefore replaced the natural promoter of *rpoBC* of the *E. coli* strain BW25113 by an IPTG-inducible promoter. The transcription of *rpoBC* was completely isolated from the upstream ribosomal proteins by introducing a selection cassette and a strong transcrip-

Reserved for Publication Footnotes

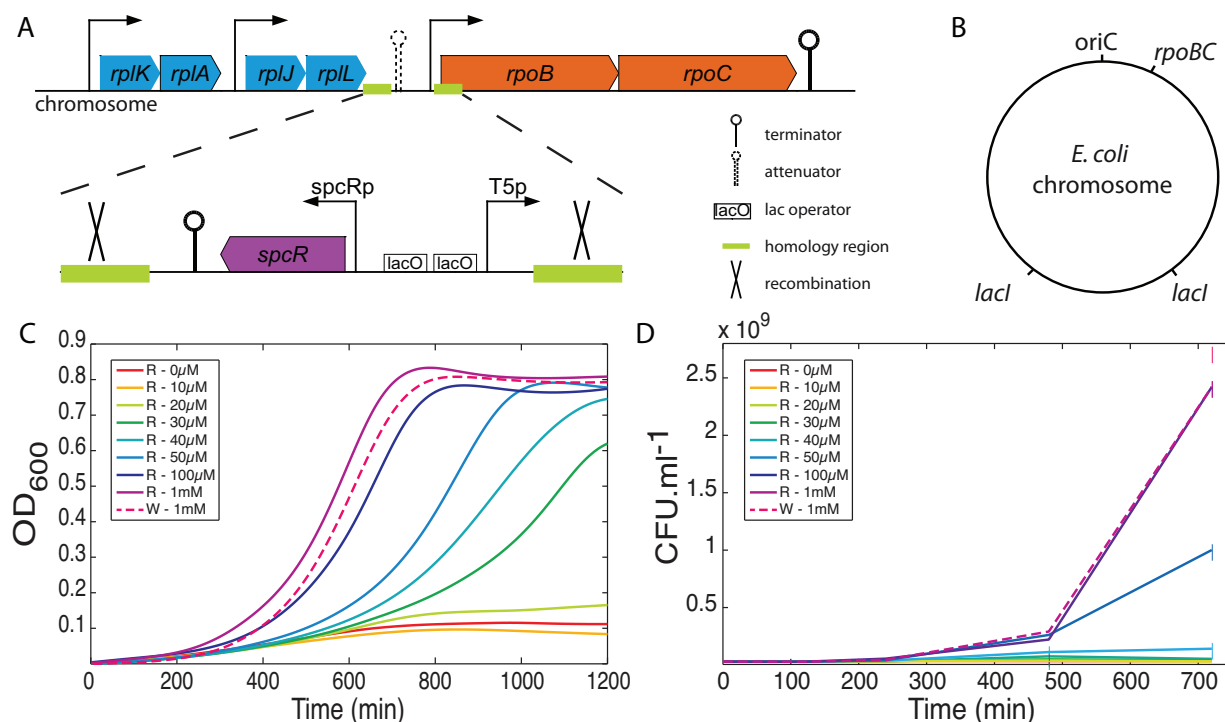


Fig. 1. RNA polymerase concentration controls the growth rate of *E. coli*. We have replaced the *rpoBC* promoter region by an inducible promoter. A strong transcriptional terminator, *rrnBt1*, was inserted upstream the T5 promoter and the selection cassette (*spcR*). *rpoBC* transcription is thus repressed by Lac repressor (A), present in two additional copies on the chromosome (B). The growth rate of the engineered strain, R, changes as a function of the concentration of IPTG added to the medium. Growth can be measured by absorbance (C) or colony forming units, CFU, (D). Whereas the absorbance of the culture keeps increasing slightly even in the absence of IPTG, the number of cells does no longer increase (red curve in C and D).

tional terminator (see Fig. 1A-B and Sec. S1 for the details of the construction). Two extra copies of the Lac repressor were added at two different loci to the chromosome in order to avoid that mutations that inactivate *lacI* relieve repression. Whereas a bacterial population (we measure typically between 10^6 and 10^9 bacteria) most likely contains an individual carrying a mutation in one copy of *lacI*, a double mutation that inactivates both copies of the gene is very unlikely. Similar attempts to control RNA polymerase levels in *E. coli* were mentioned in a conference manuscript by Nomura, but no such study has been published elsewhere (18).

RNA polymerase concentration controls growth rate. We tested the functioning of our growth controller by monitoring the growth kinetics of the engineered strain (named "R" thereafter) in a 96-well microplate (Fig. 1C) and a classical batch culture in an Erlenmeyer flask (Fig. S3). Bacteria from the different strains were grown overnight in minimum glucose medium, supplemented with a high concentration of IPTG (1 mM) to ensure maximal production of RNA polymerase. The overnight cultures were washed and resuspended in fresh minimal glucose medium containing different concentrations of IPTG. The kinetics of growth was monitored by measuring the optical density at 600 nm (Fig. 1C) and compared to an equivalent culture of the wild-type strain (named "W" thereafter). The concentration of IPTG controls growth of the R strain in a dose-dependent manner. Above 100 μM of IPTG, R cultures grow at about the same maximal growth rate as the W strain. These cultures enter stationary phase at around 0.8 OD₆₀₀. Below 20 μM of IPTG, bacterial growth is severely affected: R cultures no longer grow exponentially and reach an OD₆₀₀ of at most 0.16. Between 30 and 100 μM the growth rate of R cultures increases gradually from 0.19 h⁻¹ to 0.4 h⁻¹. The cultures reach a final OD₆₀₀ between 0.61 and 0.77, close to the value of the wt bac-

teria. We removed samples of the bacterial culture at mid-exponential phase and determined the relative intracellular concentration of the β and β' subunits by Western blotting (Fig. 1). At this stage of growth, the fully induced R strain (1 mM IPTG) and the W strain contain the same concentration of the β' subunit of RNA polymerase. Decreasing the IPTG concentration to 30 μM reduces the intracellular level of β' three-fold and the subunit is barely detectable in the absence of IPTG. These low intracellular concentrations of the two large subunits of RNA polymerase severely limit growth of the R strain (Fig. 1C).

The growth curves are identical for all concentrations of IPTG at the beginning of the experiment, suggesting that there is still enough RNA polymerase from the overnight pre-culture to ensure a maximal growth rate (Fig. 1C). Induction of the *rpoBC* operon by low concentrations of IPTG (below 100 μM) does not sufficiently stimulate the production of RNA polymerase to compensate degradation and dilution due to cell division. The steady-state concentration of the core enzyme decreases to levels too low to support maximal growth. These results demonstrate the effectiveness of our growth controller: limiting the intracellular concentration of $\beta\beta'$ is sufficient to control the growth rate of *E. coli*.

A limiting amount of RNA polymerase blocks cell division. What are the physiological changes of the cell when a low RNA polymerase concentration limits growth? Even at zero IPTG, we observe and increase in the OD₆₀₀ of the culture (Fig. 1C), but cell counting shows that the number of cells remains constant (Fig. 1D). The measurement of the optical density depends on the number of cells and their shape and size. The increase of the OD₆₀₀ during the first growth phase thus reflects a modification of cell morphology rather than cellular growth. Limiting the concentration of RNA polymerase inhibits

cell division of *E. coli* and this block of cell division is complete when the overnight culture is grown in the absence of IPTG, and therefore without any new synthesis of RNA polymerase β and β' subunits.

The change in cell morphology can be directly observed by microscopy (Fig. 3A). Time zero of this experiment is the same as in Fig. 1C-D: cells were grown overnight in the presence of high levels of IPTG, then transferred to a growth medium containing different IPTG concentrations. The initial cell population are therefore stationary phase cells just after dilution into fresh growth medium. With 1 mM of added IPTG, cells of both W and R strains have comparable sizes and shapes, the cells divide normally (time points between 0 and 8 hours), and eventually reach stationary phase (time point at 24 h). At this points, the cells in both strains are small, almost spherical, and they start to aggregate. These microscopy results are consistent with the growth kinetics shown in Fig. 1C-D.

In the absence of IPTG, the R strain does not divide between 2 and 8 hours after inoculation. No septum formation could be observed and cells form long filaments. At 24 h, R cells are about twenty times longer than W cells and the longest filaments reach 40 μm . The drastic change of cell morphology explains the moderate increase of optical density and the constancy of colony forming units observed for this condition in Fig. 1C and D. At 30 μM IPTG we observe a mixture of normal and filamentous cells at 8 and 24 hours (Fig. 3A). These results show that the immediate response of *E. coli* to low concentrations of RNA polymerase is the arrest of cell division and filamentation.

DNA replication is active in growth-arrested cells. Filamentation is a hallmark of the stress responses of bacteria, elicited, for instance, in the case of the SOS response following DNA damage (19) or during the attempt of pathogenic bacteria to evade the immune response of the host (20). When cell division is blocked, but most other physiological process remain functional, the filamentous cells still replicate their chromosome and separate their nucleoids. Fluorescent imaging of DNA stained with the Hoescht fluorescent dye reveals that filamentous R cells contain several well separated nucleoids (Fig. 3B).

Filaments ten times longer than normal W cells typically contain 10 nucleoids. The filamentous form of the R strain thus resembles normal *E. coli* cells blocked at cell division. The decreased concentration of RNA polymerase thus seems to block cell division, but not DNA replication or cell elongation.

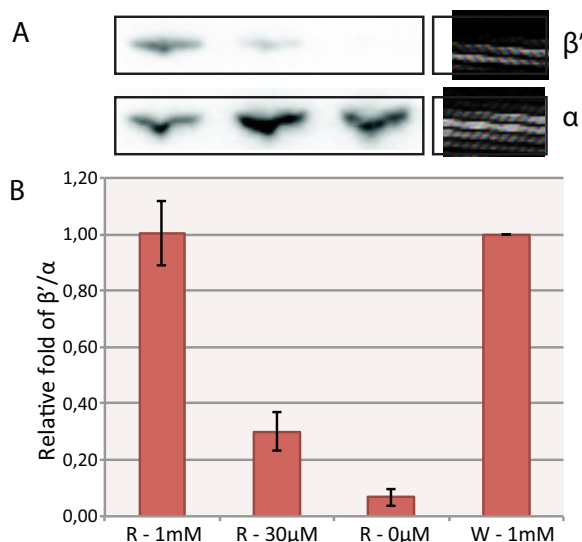


Fig. 2. Control of expression of the β' subunit of RNA polymerase. (A) Western blot of RNA polymerase subunits in the R-strain in mid-exponential phase (5h of growth in Fig. S3). Equal amounts of total protein were loaded in each lane and the band intensities were quantified. (B) Ratio of β'/α . The histogram shows the concentration of β' normalized to the concentration of α . Error bars show the standard deviation of four independent experiments. Maximal induction of the promoter (1mM IPTG) in the R-strain produces the same concentration of the β' subunit as the W strain.

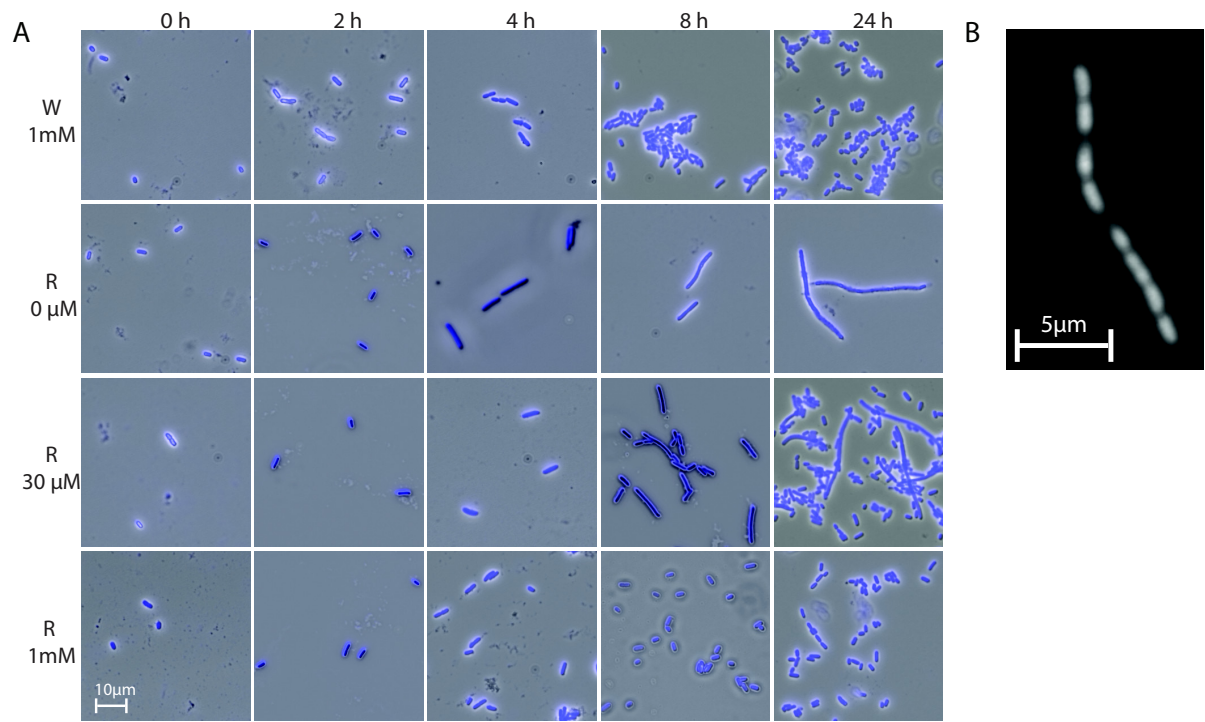


Fig. 3. Size and shape of cells as a function of *rpoBC* expression. (A) Cells were sampled (Fig. S3) at the times indicated at the top of each column, fixed and visualized by fluorescence (blue) and phase contrast microscopy (shades of gray). The concentration of IPTG added to the medium is shown at the left of each row. (B) An enlarged view of a fluorescence image of one cell after 16 h of growth in the absence of IPTG. Multiple nucleoids can be distinguished in the elongated cell.

Growth arrest is reversible. Bacterial stress responses to environmental threats are often reversible. For example, bacteria can recover from filamentation when moved to favorable environmental conditions (21). To see whether the growth arrested, filamentous cells can recover a growth phenotype, we monitored the kinetics of growth of the R strain at the single cell level using a microfluidics device. The device traps cells in a channel about 1.5 times the diameter of a bacterial cell and about 20-times as long as an exponentially growing cell. The medium flows from the top of the channel (wide opening) to the bottom and thus traps bacteria in the channel. The constant flow of medium assures constant growth conditions and allows an easy change of growth medium. In our case, the media change consists in adding or removing IPTG from the M9 minimal glucose medium. Growth of the bacteria was observed by time-lapse microscopy to obtain phase contrast and fluorescent images (see Fig. 4 and supplementary video).

R cells trapped in the dead-end channels were initially grown in the presence of 1 mM IPTG for 13 hours in minimal glucose medium. The synthesis of RNA polymerase is strong and the cells have a normal cell cycle. They have the usual cylindrical shape; they grow and divide until they completely fill the dead-end channel (Fig. 4a). New-born cells are then washed out into the large channel that supplies the medium. After 13.5 hours of growth, IPTG was removed from the

medium. Two hours later, cells start to elongate, growing to a length of about 20 micrometer after 6 hours without IPTG (Fig. 4b). At this point, we again added IPTG to the growth medium. One hour after this new buffer change 90% of the 950 tracked cells start dividing again (Fig. 4c) and normal growth resumes (Fig. 4d). These results show that filamentation caused by low RNA polymerase levels is a reversible process. Cell division takes up again when the intracellular concentration of RNA polymerase rises.

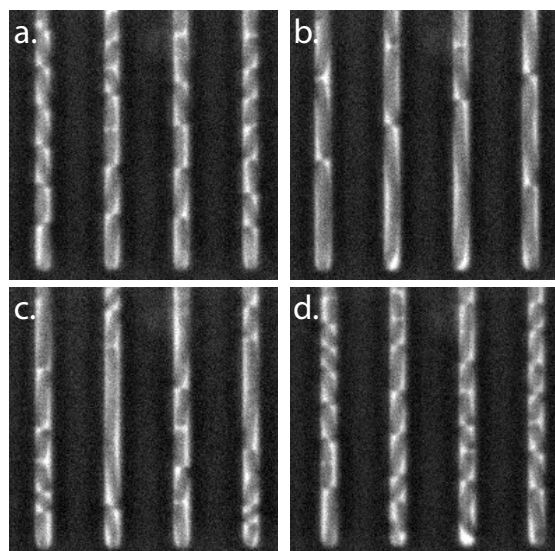


Fig. 4. Reversible growth arrest and re-start. Bacterial strain R was grown in a microfluidics device and phase contrast images were acquired every 10 minutes. The cells were grown initially in the presence of 1 mM IPTG (A). Six hours after removing IPTG from the medium, the cells are very elongated (B). One hour after adding back 1 mM IPTG into the medium, the cells divide (C), and resume normal growth (D). See supplementary movie 1 for the kinetic details.

***E. coli* compensates for decreased RNA polymerase concentration by an increased transcription of ribosomal proteins.** Cell division is thus clearly affected by low concentrations of RNA polymerase. Since RNA polymerase is responsible for transcription, we compared the transcriptome of growth-limited R cells to the expression profile of the W strain. Both strains were pre-cultured overnight in minimal glucose medium supplemented with 1 mM IPTG and transferred to fresh media containing various concentrations of IPTG. RNA sequencing was performed on mRNAs isolated from cells harvested after 4 to 5 hours of growth (Material and Methods). This time point corresponds roughly to the sample time for the Western blot (Fig. 1).

In order to obtain an overall picture of the transcriptional changes at different concentrations of RNA polymerase, we measured the expression profile in four conditions: W (W strain), R0 (R strain without IPTG in the growth medium), R30 (R-strain grown with 30 μ M IPTG), and R1000 (R-strain grown in the presence of 1 mM IPTG). Equal amounts of total RNA were prepared from these samples and used for RNAseq. Since a precise determination of the RNA content per cell proved to be difficult, we assumed in the subsequent analysis that the total amount of RNA per cell was identical in the different conditions. We therefore report relative expression changes, which are independent of this assumption.

We compared the expression profiles using an unsupervised multivariate factorization technique, the correspondence analysis (COA; section Sxx). This method has been extensively used to analyze microarray data (22) and, more recently, for RNA-seq data (23). This analysis method measures the overall resemblance of the expression profiles and locates each expression profile in a plane. Samples with similar expression profiles are spatially close. Our data sets are thus separated into three distinct groups: R1000, W, and R0/R30 (Fig. 5A). R0 and R30 are very close and can be considered a single group. Surprisingly, most of the genes are clustered at equal distance between the three groups, which indicates that there are no major dif-

ferences in gene expression when the intracellular RNA polymerase concentration is varied (see also box plots in Fig. S7).

We use the TMM method implemented in edgeR to identify differentially expressed genes. We considered genes to show significant differential expression when the \log_2 expression ratio is above 2 or below -2; in other words, the change in gene expression is at least four-fold. Only a subset of genes at the periphery of the cluster satisfies this criterion (colored in Fig. 5A). As expected, the two genes that we control externally, *rpoB* and *rpoC*, are well set apart from the cluster. Their close proximity with the R1000 group reveals an overexpression of these genes in the R1000 condition compared to W: their \log_2 ratios comparing R1000 to W are 3 and 1.8 for *rpoB* and *rpoC*, respectively. Even though transcription *rpoBC* is stronger in R1000 compared to W, the concentration of the corresponding proteins is very similar in these two strains (Fig. 1). Previous experiments have shown that over-expression of *rpoBC* can be compensated by translational repression of these genes (17). As expected, these same two genes are greatly under-expressed in R0 and R30 with respect to W and R1000, with \log_2 ratios between -3.7 and -4.4.

To identify specific cellular functions that are differentially affected by changes in *rpoBC* expression, we classified the whose expression was significantly altered (\log_2 ratios above 2 and below -2) into seven categories based on their multifunctional classification (see Materials and Methods). The pie charts in Fig. 5B show that no particular functional category is particularly sensitive to variations of the RNA polymerase concentration: the relative proportion of the different sectors of the pie-chart remain essentially constant in the different comparisons. Furthermore, we did not observe any very large changes in the level of gene expression, which is consistent with the cluster of genes observed with the COA analysis. The biggest number of genes with significant expression changes occurs in the R0 strain compared to the W strain, with 620 genes changing expression by more than a factor of four (Fig. 5B).

Although very few genes are differentially expressed in R1000, the ribosomal protein genes are a striking exception. Compared to wild type, they tend to be under-expressed in this condition (Fig. 5C and S7). Some of them are even over-expressed in R0 and R30. Remember that the expression ratios should be interpreted in relative terms, since the absolute quantification of total mRNA in the different cells is rather error prone. Independently of this caveat, these data clearly show that a decreased concentration of RNA polymerase in the R0 and R30 conditions is partially compensated by an increased expression of the ribosomal proteins, i.e., of the translation machinery. Conversely, an increase in the concentration of RNA polymerase in the R1000 conditions leads to a decreased expression of ribosomal proteins.

Discussion

Growth rate is probably the most important physiological parameter of a bacterial cell. The competition between different strains or mutants is based on achieving an optimal growth rate for a given environmental condition. Growth rate regulation is thus generally thought of as the optimal resource allocation for a given, limiting nutrient. The cell adjusts gene expression and enzyme activities to meet external constraints. In this manuscript we have investigated growth rate control from the opposite perspective by setting the growth rate not through the availability of a nutrient source, but through the availability of an essential cellular component: RNA polymerase.

We have engineered a strain where the large subunits of RNA polymerase, β and β' , are transcribed from an IPTG-inducible promoter. In this strain, growth rate can be tuned by externally setting the concentration of RNA polymerase. As expected, overproducing the large subunits of RNA polymerase does not increase growth rate. This result shows that RNA polymerase is not the factor limiting growth rate. This result is entirely consistent with previous observations of the effects of overexpressing the *rpoBC* genes (24). Interme-

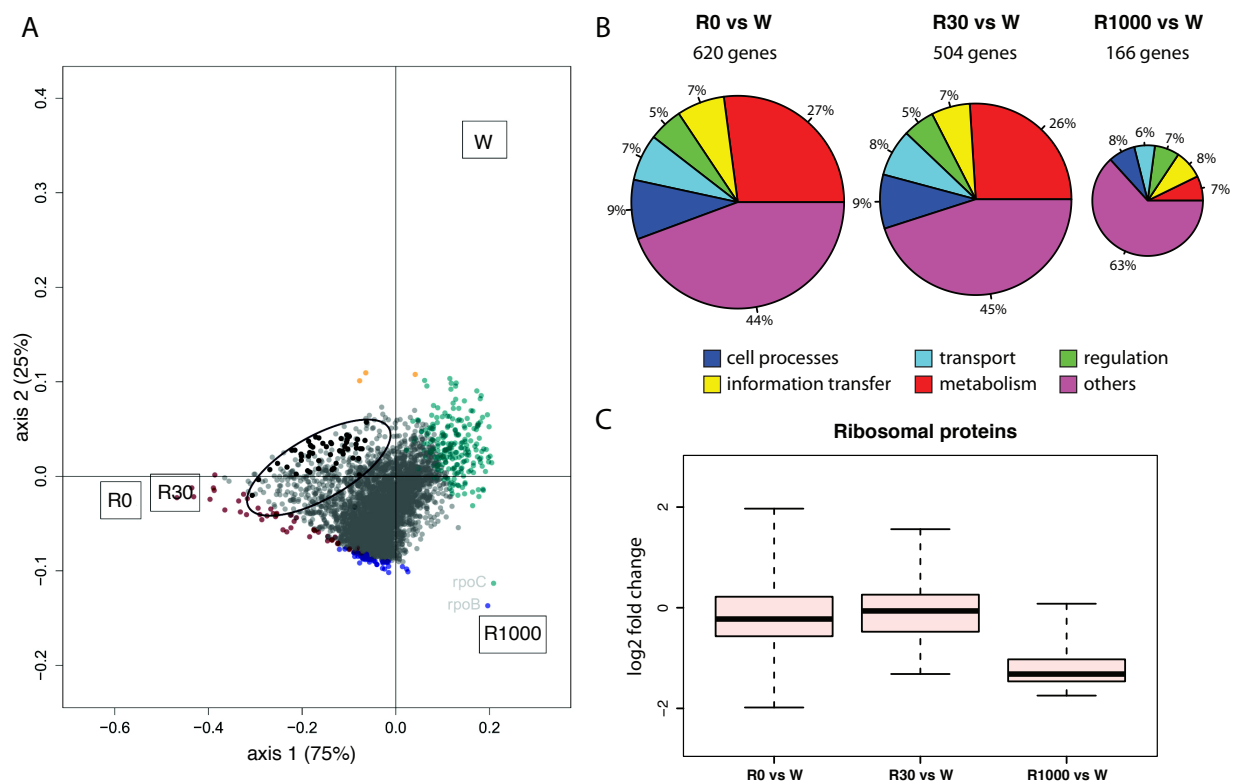


Fig. 5. Analysis of the transcriptome obtained by RNAseq. (A) COA analysis of the W and R strains at various IPTG concentrations. 98% of the data are explained by the two first dimensions. The x-axis separates 73% of the data and the y-axis account for almost all of the remaining differences (25%). Each dot corresponds to a gene. The color code distinguishes genes whose expression is significantly different compared to the W, as determined by edgeR analysis. Genes in green correspond to highly-repressed genes in R0 with respect to W (\log_2 ratio < -2), while genes colored red represent highly-expressed genes in R0 compared to W (\log_2 ratio > 2). Genes in blue are highly expressed in R1000 compared to W, and the ones colored orange are under-expressed in R1000. Genes in gray are not significantly over- or under-expressed in the different conditions (\log_2 ratios between -2 and 2). Circled genes correspond to ribosomal protein genes, which are under-expressed in R1000 with \log_2 ratios of -1 , on average. Some of these genes coding for ribosomal proteins are over-expressed in R0 and R30. (B) Pie chart of the truly differentially-expressed genes in the R and W strains. Circle areas are proportional to the number of genes differentially expressed between conditions. (C) The ribosomal protein \log_2 expression ratio for R cells compared to W are given on y-axis.

diate concentrations of RNA polymerase lead to slower growth rates and at very low concentrations of RNA polymerase the cells stop to divide. However, the metabolic activity of the cell persists, replication of the chromosome continues and biomass increases, albeit at a much lower, linear rate, by simple cell elongation. This filamentation process is entirely reversible when RNA polymerase is again produced.

Since *E. coli* promoters can have very different K_m values for RNA polymerase binding, we had expected to observe very distinct effects of decreasing the concentration of RNA polymerase on specific classes of genes. Our transcriptome analysis does not reveal such a trend. Rather, all functional classes of genes are roughly equally affected by varying concentrations of RNA polymerase and the number of significantly affected genes is relatively small (Fig. 5). The transcriptional regulatory network thus seems to be rather robust to perturbations of the key component of the network. Part of this robustness may be due to a partial compensation of inefficient transcription by increased translation. The transcription of the ribosomal proteins is inversely correlated with the concentration of RNA polymerase. The molecular mechanism of this regulation is not clear. Our analysis of RNAseq data has filtered unique mappings and did therefore not quantify ribosomal RNA. Since the transcription of ribosomal proteins is regulated by the availability of ribosomal RNA, it is likely that the observed regulatory effect is exerted at the level of transcription of ribosomal RNA. Indeed, the seven ribosomal pro-

motors of *E. coli* are among the strongest promoters of the bacterium, possessing the lowest K_d values of binding RNA polymerase. Low concentrations of RNA polymerase could therefore selectively favor the transcription of *rrn* genes and thus, indirectly, the transcription of genes coding for ribosomal proteins. This could represent a very general compensatory mechanism between transcription and translation.

The analysis of the transcriptome did not reveal the molecular reason why low concentrations of RNA polymerase induce filamentation, a rather common response to different stress situations such as DNA damage. Genes of the SOS response are not particularly induced. Even though the *fts* genes are weakly expressed at low polymerase concentrations, their expression profile is not significantly different from most average genes of the cell. Irrespective of the molecular mechanism underlying filamentation, the physiology of these division arrested cells could be exploited in a biotechnological applications. Since the cells containing low concentrations of RNA polymerase are still metabolically active but convert few of the available nutrient resources into biomass, they could potentially be used for the efficient production of high value-added metabolites.

Materials and Methods

Bacterial strains.

The strains are derivatives of *E. coli* K-12 BW25113 strain. Our reference strain IJ39 (W) carries two copies of the *lacI* gene to provide a large amount of the lac repressor and to prevent the appearance of fixed mutation. In addition to these extra copies of *lacI*, our modified strain IJ40 (R) carries on its chromosome a IPTG-inducible promoter in front of the *rpoBC* operon.

Growth conditions and measurements.

Cells were grown at 37°C in minimal M9 medium supplemented with 0.2% of glucose. For all measurements, the cultures were inoculated from an overnight culture, grown for 18 hours in an equivalent medium containing 1 mM IPTG. The growth kinetics were measured either in an Erlenmeyer flask (see Fig. S3) or in microplate. At time zero, IPTG was removed from the overnight culture by centrifugation for 5 min at 4,000 g and the cells were washed with fresh M9 medium without IPTG. This operation was repeated twice and the inoculum size was adjusted to give equal optical density to all cultures. To start the kinetic experiment, the washed cultures of equal optical density were diluted 100-fold into fresh medium with or without IPTG to reach an initial OD₆₀₀ of 0.01. For cultures in microplates, we added a 2 mm glass bead to each well in order to improve culture-shaking, as well as 60 µl of mineral oil to avoid evaporation. Cultures were shaken every 2 min and the absorbance at 600 nm was read every 5 minutes by an automate plate reader (Perkin fusion FT Alpha).

Western blotting.

After 4 to 5 h of growth in M9 medium supplemented with 0.2% glucose, cells were harvested by centrifugation at 10,000 g for 10 min at 4°C. The cells were stored at -80°C and lysed using the BugBuster Master Mix solution (Novagen) supplemented by a complete mini protease inhibitor cocktail (Roche) for 20 min at room temperature. The soluble cell extracts were collected by centrifugation at 14,000 g for 20 min at 4°C and the amount of total proteins was measured using a Bradford assay (Biorad). In order to detect the α and β' subunits of RNA polymerase by Western blotting, 25 to 35 µg of total proteins were separated on a 4%-12% NuPAGE gel (Invitrogen), transferred to Hybond-P membranes (Amersham), blocked with non-fat dried milk, and incubated with 1/20,000 mouse anti- α (sc-101597 4RA2) and 1/5,000 anti- β' (sc-101613 NT73) for 1 h at room temperature. Membranes were washed and incubated another 1 h with secondary anti-mouse antibody conjugated with peroxidase. The Amersham ECL chemiluminescent kit was used to detect the signals, and Amersham Hyperfilm ECL films to visualize the results.

Microscopy and Microfluidics assays.

To visualize the nucleotide in elongated cells by microscopy, cells were fixed with a 10% formaldehyde solution at room temperature for 20 min. The cell were then pelleted by centrifugation 10,000 g for 2 min and re-suspended in absolute ethanol. An aqueous solution of Hoechst was added to a final concentration of 10 µg/ml.

Microfluidics experiments.

For the microfluidics assays, cultures were grown overnight at 37°C in M9 supplemented with 0.2% glucose and 1 mM of IPTG. The next morning, the overnight culture was diluted 50-fold into 10 ml of fresh medium at 37°C. When the culture reached OD₆₀₀ = 0.15, the cells were concentrated by centrifugation and injected into the device similar to the mother machine described in (25). The cells were loaded by centrifugation of the device, until more than 80% of the channels were filled with bacteria. We then passed fresh M9 medium supplemented with 1 mM IPTG over the cells and microscope acquisition started for about 2 days. After 13 h, the syringe containing the medium was switched to the same medium without IPTG. Finally, 6 hours later, a new syringe containing again 1 mM IPTG was plugged into the device. Pictures were taken every 10 min for 40 fields, representing a total of about 1000 individual channels.

RNA sequencing.

RNA was extracted from mid-exponential phase cultures of R and W strains using the RNeasy Mini Kit (Qiagen) and enriched in mRNAs using the MICROB-Express Bacterial RNA enrichment kit (Ambion). The integrity of RNA and the efficiency of rRNA depletion were assessed by a Bioanalyzer analysis (Agilent; Figure S5). Library construction and Illumina sequencing with a read length of 100 pb were performed by Otogenetics. Sequences were filtered for noncoding RNAs and analyzed with the DNAnexus platform using the *E. coli* MG1655 genome as reference (see Table S2 for the statistics). Genes with read counts below 10 in all conditions were removed from the data sets. Read counts were

normalized by the sequencing depth using the trimmed mean of M values (TMM) normalization method implemented in edgeR (R, Bioconductor) (26). Similar results were obtained with the DESeq normalization method (27). Differential gene expression is expressed as the log₂ expression change between R and W cultures. The genes with the greatest differential expression ratios identified by the two normalization methods were confirmed by correspondence analysis performed on the raw read counts for the four conditions (ade4 package, R). Multifunctional classification of the expressed genes was achieved with the MultiFun schema accessed through the EcoCyc database and home-made scripts (28, 29).

ACKNOWLEDGMENTS. This work was partially supported by

References

- Sonderegger M, Schümperli M, Sauer U (2005) Selection of quiescent *Escherichia coli* with high metabolic activity. *Metabolic engineering* 7:4–9.
- Matin A, Little CD, Fraley CD, Keyhan M (1995) Use of starvation promoters to limit growth and selectively enrich expression of trichloroethylene- and phenol-transforming activity in recombinant *Escherichia coli* [corrected] [published erratum appears in Appl Environ Microbiol 1995 Nov;61(11):4140]. *Applied and environmental microbiology* 61:3323–3328.
- Chotani G, et al. (2000) The commercial production of chemicals using pathway engineering. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* 1543:434–455.
- Bremer H, Dennis PP (1996) in *Escherichia coli and Salmonella*, eds Neidhardt FC, et al. (ASM Press, Washington, D.C) Vol. 2, Asm press edition, pp 1553–1569.
- Maaloe O, Kjeldgaard NO (1966) *Control of macromolecular synthesis: a study of DNA, RNA, and protein synthesis in bacteria*. Microbial and molecular biology series (W. A. Benjamin).
- Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science* 330:1099–1102.
- Schaechter M, Maaloe O, Kjeldgaard NO (1958) Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of *Salmonella typhimurium*. *Microbiology* 19:592–606.
- Rowe DCD, Summers DK (1999) The Quiescent-Cell Expression System for Protein Synthesis in *Escherichia coli*. *Appl. Environ. Microbiol.* 65:2710–2715.
- Alper H, Stephanopoulos G (2007) Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metabolic engineering* 9:258–67.
- Engbaek F, Gross C, Burgess RR (1976) Biosynthesis of *Escherichia coli* RNA polymerase subunits upon release of rifampicin inhibition. *Molecular & general genetics : MGG* 143:297–9.
- Kawakami K, Saitoh T, Ishihama A (1979) Biosynthesis of RNA polymerase in *Escherichia coli*. IX. Growth-dependent variations in the synthesis rate, content and distribution of RNA polymerase. *Mol Gen Genet* 174:107–116.
- Shepherd N, Dennis P, Bremer H (2001) Cytoplasmic RNA Polymerase in *Escherichia coli*. *Journal of bacteriology* 183:2527–34.
- Yamamoto M, Nomura M (1978) Contrascripton of genes for RNA polymerase subunits beta and beta' with genes for ribosomal proteins in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 75:3891–5.
- Dykxhoorn DM, Pierre RS, Linn T, St Pierre R (1996) Synthesis of the beta and beta' subunits of *Escherichia coli* RNA polymerase is autogenously regulated in vivo by both transcriptional and translational mechanisms. *Mol Microbiol* 19:483–493.
- Ishihama A, Fukuda R (1980) Autogenous and post-transcriptional regulation of RNA polymerase synthesis. *Mol*

- Cell Biochem* 31:177–196.
16. Campbell Ea, et al. (2001) Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* 104:901–12.
 17. Hayward RS, Fyfe S (1978) Over-synthesis and instability of sigma protein in a merodiploid strain of Escherichia coli. *Molecular & general genetics : MGG* 159:89–99.
 18. Nomura M, Bedwell DM, Yamagishi M, Cole JR, Kolb JM (1987) *RNA polymerase and regulation of rRNA synthesis in Escherichia coli: RNA polymerase concentration, stringent control, and ribosome feedback regulation* eds Reznikoff WS, et al. (Elsevier, New York), pp 137–149.
 19. Walker GC (1996) in *Escherichia coli and Salmonella : cellular and molecular biology*, Progress in Molecular Biology and Translational Science, eds Neidhardt F, et al. (ASM Press) Vol. 1, 2nd edition, pp 1400–1416.
 20. Justice SS, Hunstad DA, Cegelski L, Hultgren SJ (2008) Morphological plasticity as a bacterial survival strategy. *Nature reviews. Microbiology* 6:162–8.
 21. Young KD (2010) Bacterial shape: two-dimensional questions and possibilities. *Annual review of microbiology* 64:223–40.
 22. Hoheisel D, et al. (2001) Correspondence analysis applied to microarray data. 98.
 23. Pierron G, et al. (2012) A new subtype of bone sarcoma defined by BCOR-CCNB3 gene fusion. *Nature genetics* 44:461–6.
 24. Guzman EC, Jimenez-Sanchez A (1986) Effect of multiple copies of rpoBC on the rate of RNA synthesis in Escherichia coli. *Genetical research* 48:61–4.
 25. Wang P, et al. (2010) Robust growth of Escherichia coli. *Current biology : CB* 20:1099–103.
 26. Robinson M, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.*
 27. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome biology* 11:R106.
 28. Serres MH, Goswami S, Riley M (2004) GenProtEC: an updated and improved analysis of functions of Escherichia coli K-12 proteins. *Nucleic acids research* 32:D300–2.
 29. Keseler IM, et al. (2011) EcoCyc: a comprehensive database of Escherichia coli biology. *Nucleic acids research* 39:D583–90.

Supplementary information

To

Growth control of *Escherichia coli* by the global transcriptional machinery

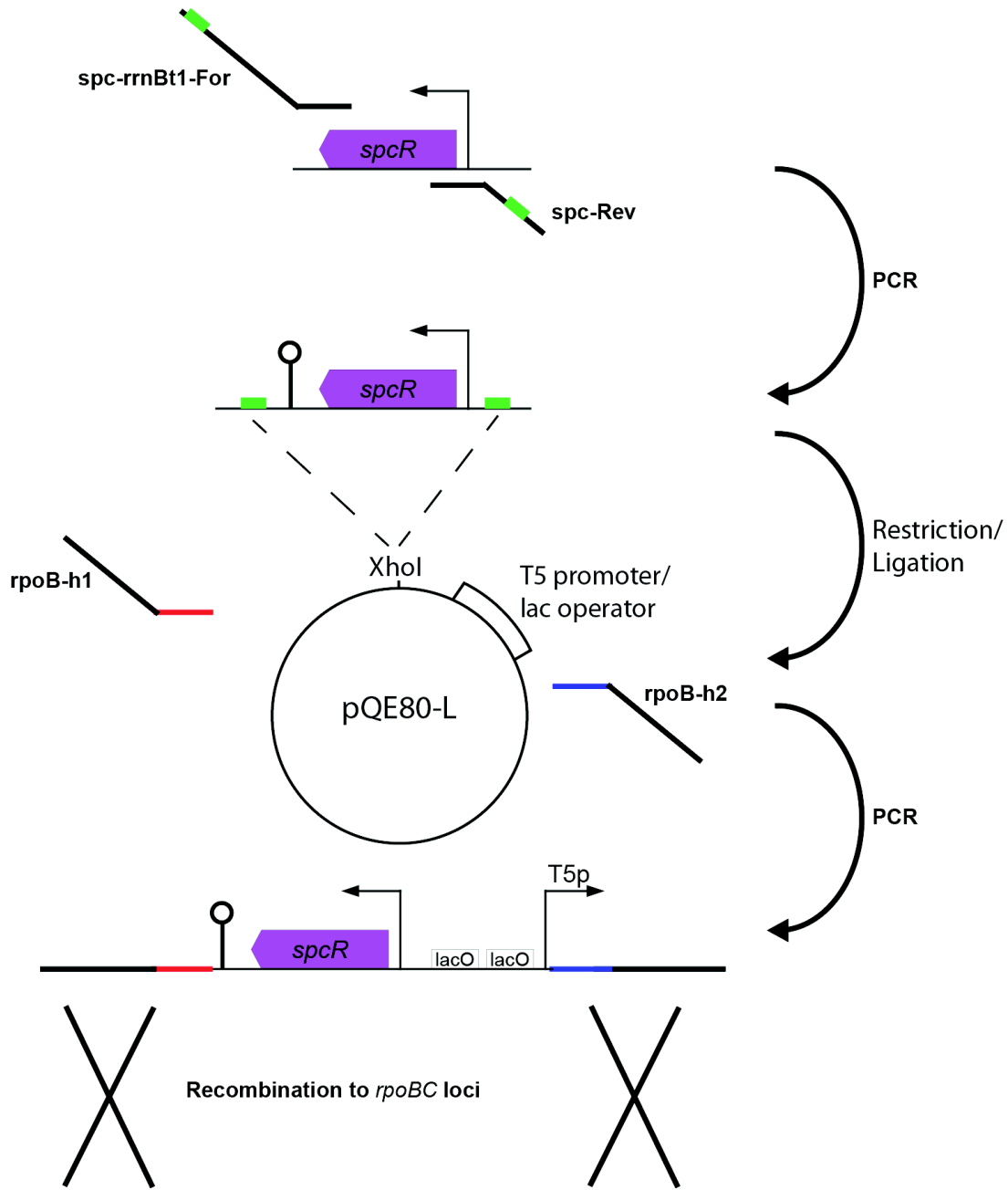


Figure S1: Construction of the inducible *rpoBC* operon

A spectinomycin-resistance (*spcR*) cassette was PCR amplified from the pGBM2 plasmid using primers *spc-rrnBt1* and *spc-Rev* (SI Table 1). They both carry a *XhoI* restriction site and a strong bidirectional terminator in the case of *spc-Rev*. We cloned this cassette upstream of the T5 *lac*-inducible promoter of the pQE80-L plasmid (Qiagen). The promoter contains two *lac* operator sequences which ensure efficient repression by the *lac* repressor. Using primers *rpoB-h1* and *rpoB-h2* (SI Table 1), we PCR amplified the region spanning the *rrnBT1* terminator, the *spcR* cassette and the T5 pQE80-L promoter. This PCR product was eventually recombined upstream of the *rpoBC* CDS (in place of the 320pb DNA region between *rplL* and *rpoBC*).

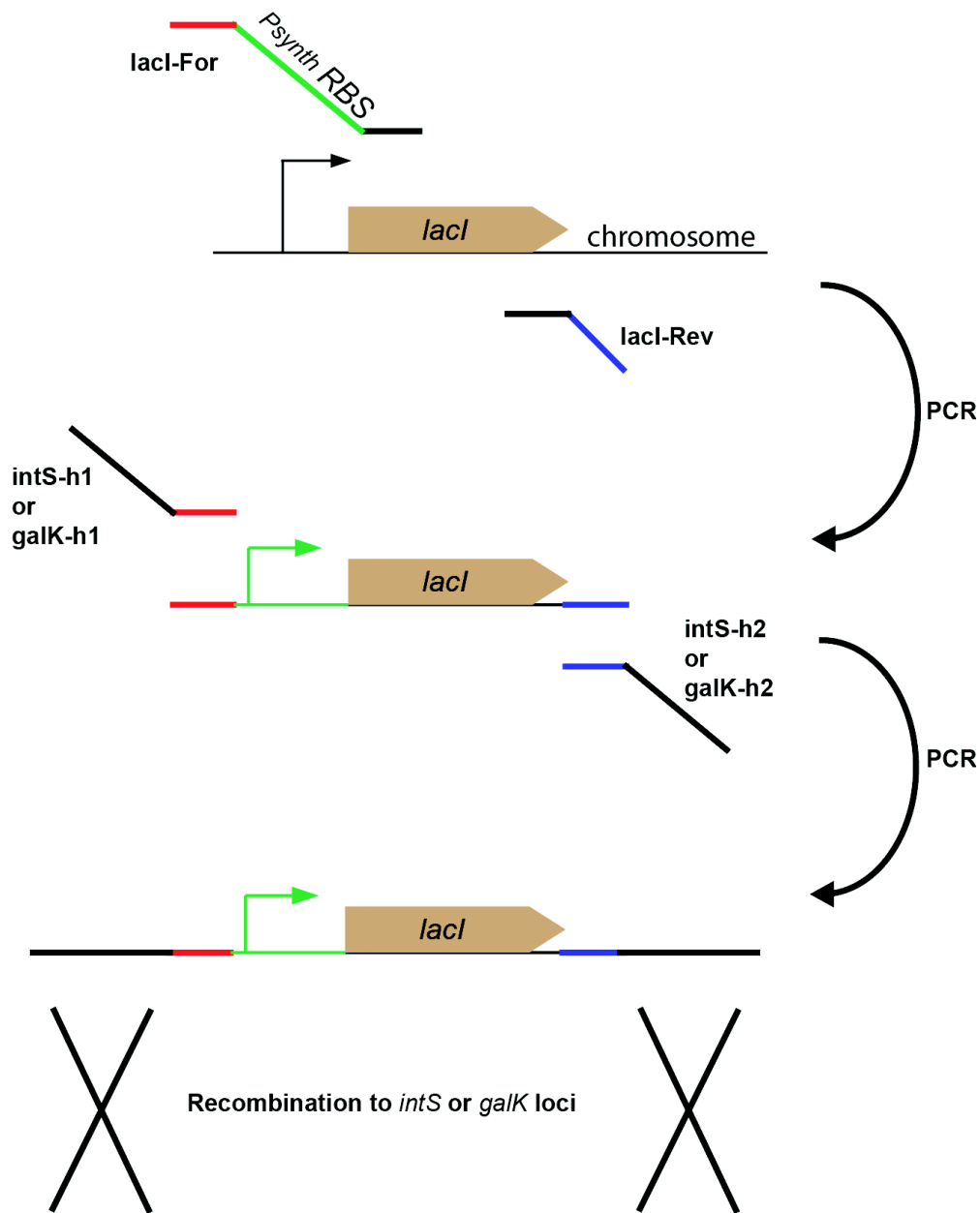


Figure S2: Construction of a strong expressed *lacI* on the chromosome in *galk* and *intS* loci.

To ensure a high expression of the *lac* repressor, we introduced a strong synthetic promoter (Braatsch et al. 2008) into the forward primer *lacI-For*. We recombine these extra copies of *lacI* in replacement of the *galk* and *intS* genes by using a *ccdB*-toxin counter-selection system (Ranquet et al., manuscript in preparation) without any antibiotic resistance cassette left on the chromosome. The sequences of the primers are listed in SI Table 1

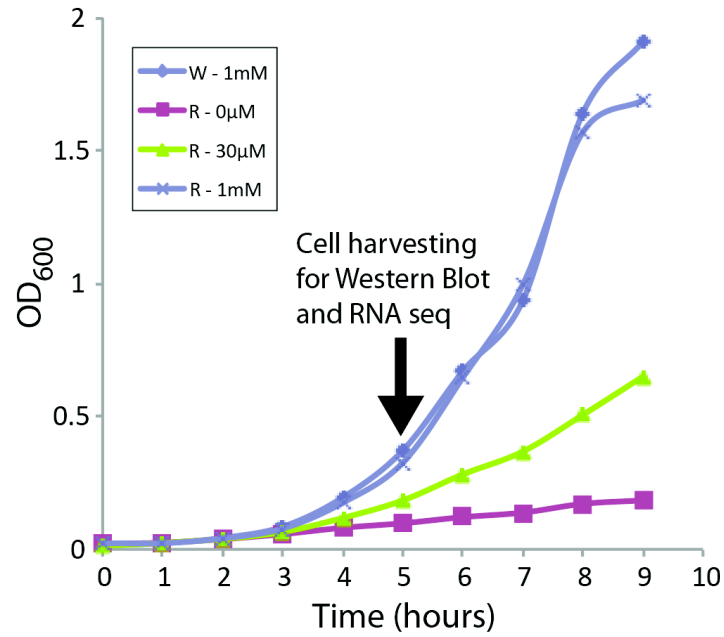


Figure S3: Dose-dependent effect of *rpoBC* gene expression on cell during growth in flask.

The growth rate of the engineered strain R increases in presence of IPTG in a dose dependent manner when cells are grown in flask as when they are grown in microplate (figure 1c). At time point 5h cells were harvested for the Western blot and RNA seq experiments (figure 3 and figure 5).

Movie.mp4

Figure S4: Effect of rpoBC gene expression on cells physiology and growth overtime, in a microfluidic device

R strain cells were trapped in the microfluidic channels so they can be monitored during long period (here 43h). The experiment starts with presence of IPTG in the medium and cells behave as the W strain meaning: they divide normally and have a normal size and shape. After 13h30, the flow of fresh medium is switched to a medium without IPTG. Two hours after this switch (15h30) the cells stop dividing and start to elongate. After 6h without IPTG in the medium they reach the maximum size and we switch back to a medium with IPTG (to avoid cells jumping out form the channel because of there size). One hour after adding IPTG back, we observed the recovery of division process for the first cells and at the end of the experiment 90% of cells recovered.

This movie shows 4 fields of the 50 we observed during a timelapse microfluidics experiments. The figure 4 in the manuscript results in some snapshot taken from the first field of this movie.

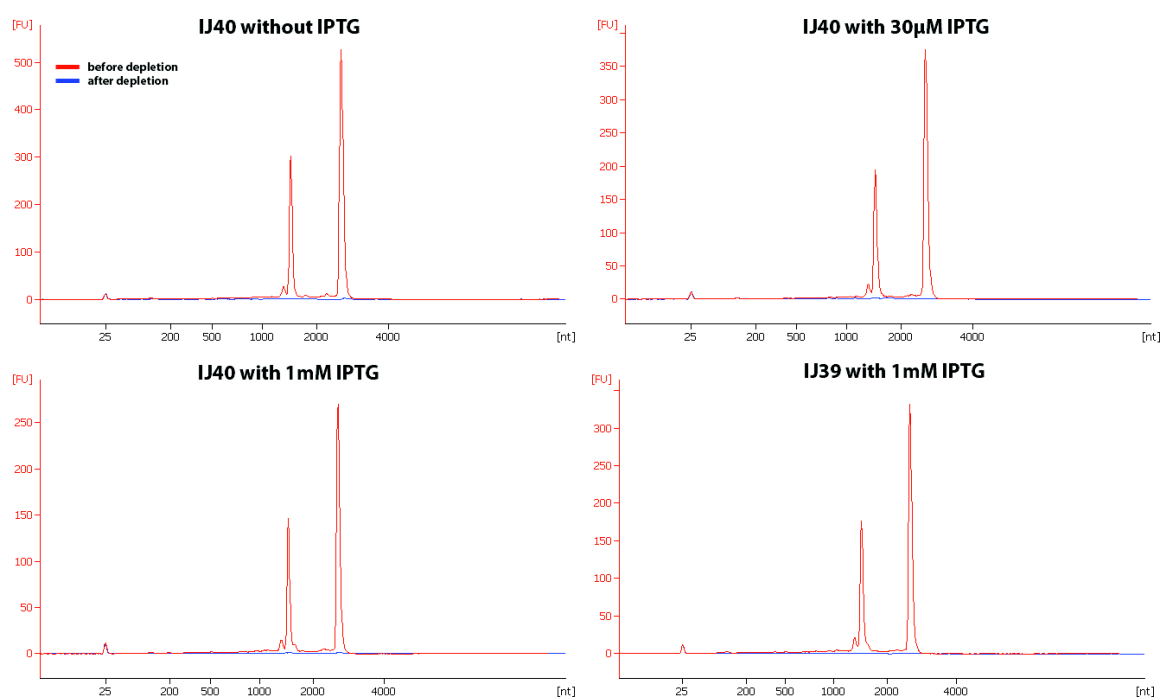


Figure S5: Quality check by Agilent of RNA used for the RNA sequencing.

Total RNA were extracted from the W strain in presence of IPTG and the R strain in tree different concentration of IPTG. For each of these RNA extracts we enriched in mRNA (using Ambion kit) to improve the RNA seq resolution. To check the quality of the sample we made an agilent assay on each sample before and after the rRNA depletion. We observed that the 16s and 23s represented between 46 and 54% of the total RNA before treatment and only 4 to 6 % after depletion.

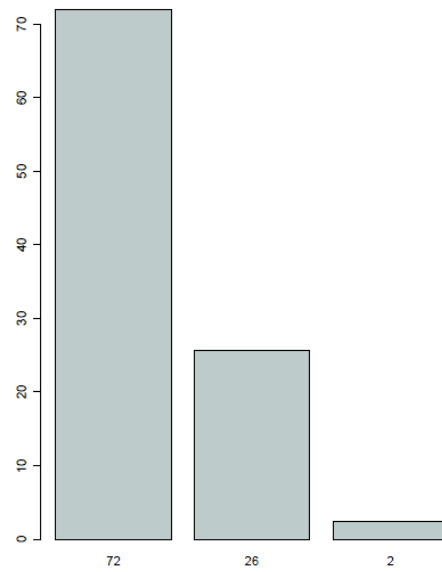


Figure S6: Bar plot for the correspondence analysis of the four conditions tested in RNA-Seq.

Data sets were separated along three dimensions. The percentage of inertia is indicated below each bar.

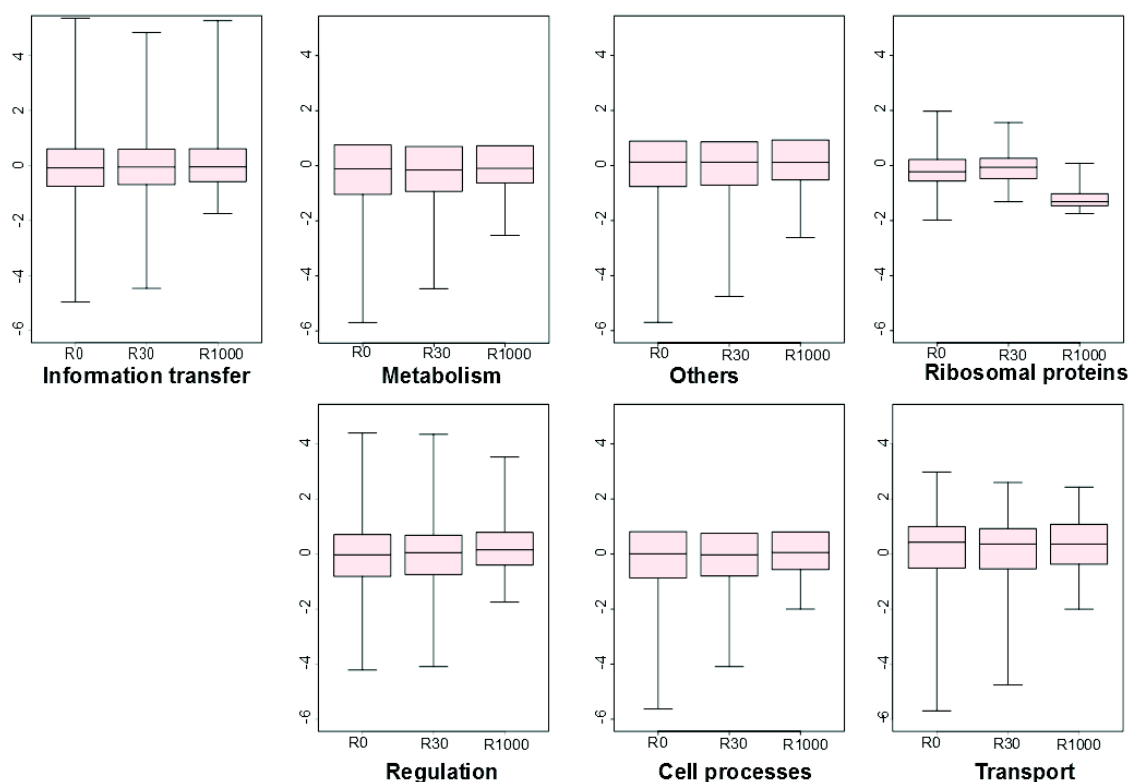


Figure S7: Functional classification of the differentially-expressed genes.

The log₂ expression ratios for R cells compared to W are given on the y-axis. Genes were classified into 7 categories according to Serres *et al.* (2004). The category “Others” groups genes from the categories “extrachromosomal”, “DNA sites”, “cryptic genes”, and genes that have no functional annotation.

Table S1: List of primers for constructions

Below are listed all the primers we used to construct the W and R strains. Highlighted nucleotides correspond to homology sequences (blue and red) and restriction sites (green) mentioned in Fig. S1 and S2

Primer name	Primer sequence
spc-rnBt1-For	ACGTCTCGAGATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCT TTCGTTTTATAATTGTTAGACATTATTTGC
spc-Rev	CTGATAGGGACTCGAGCCAG
rpoB-h1	AGCCTTTTTGCGCTGTAAGGCGCCAGTAGCGTTTCACACTGTTG ACTACTGCTGTGCCTGGCCCTTCGTCTTCACCTC
rpoB-h2	TTGTGGACGTTTACCAAATCCTTACGAATACGTTTTTCTCGGTA TAGGAGTAAACCATAGTTAATTCTCCTCTTTAA
lacI-For	TAAGCTAGCGTGACGATGCGTTGACATATCACTGTGATTCACATA TAATATGCGAAATCAGAAGAGTATTGCTAATGAAACCAGTAACG TTATA
lacI-Rev	TAAGCTAGCTGATCACCTTAGTCACTGCCCCGCTTTCCAGTC
intS_h1	CCGTAGATTACAGTTCGTCATGGTTCGCTTCAGATCGTTGACAG CCGCATAAGCTAGCGTGACGATGCG
intS_h2	ATAGTTGTTAAGGTCGCTCACTCCACCTTCTCATCAAGCCAGTCC GCCCCATAAGCTAGCTGATCACCTTAG
galK_h1	GTTTGCGCGCAGTCAGCGATATCCATTTTCGCGAATCCGGAGTGT AAGAATAAGCTAGCGTGACGATGCG
galK_h2	TTCATATTGTTTCAGCGACAGCTTGCTGTACGGCAGGCACCAGCTC TTCCGTAAGCTAGCTGATCACCTTAG

Supplementary Table 2: Analysis of the RNA-seq data mapped to the *E. coli* MG1655 genome

Strain and condition	WT	R 0	R 30	R1000
Mapped reads	21, 314, 881 (82.55%)	22 , 195 , 528 (82.71%)	19 , 133 , 573 (84.35%)	29, 569, 588 (84.57%)
Reads not mapped	17.45%	17.29%	15.65%	15.43%
Total reads	25,820,054	26,833,916	22,682,308	34,965,492
Total bases	2,582,005,400	2,693,391,600	2,268,230,800	3,496,549,200

Supplementary text

Western blot quantification

For quantification we also took images of membranes using a camera CCD amplified. Digital Western blot images were analysed using imageJ (<http://rsbweb.nih.gov/ij/>). The density of each band was measured and background-subtracted. Since the expression of the Alpha subunit is constitutive, the band-intensity for the β' subunit (155.16 kDa) was divided by the one for α (36.512 kDa) to deliver a relative expression value for Beta' in each strain at different IPTG concentrations. The quantification was made on four independent experimental datasets.

RNA-Seq: Correspondence analysis

Correspondence analysis was performed on raw data, i.e. read counts for the four conditions tested without prior normalization. The variation of the data can be explained by three dimensions (Fig. S6). The first dimension of the analysis accounts for 72% of the total variability, while the second and third dimensions account for 26% and 2%, respectively.

The plot of the two first dimensions is given in Fig. 5 of the main text. The raw read counts allow discriminating the data sets into 3 distinct groups: R0-R30, R1000, and W. Along the first dimension, R0-R30 and R1000-W are best separated, while R1000 and W are separated along the second dimension. R0 and R30 can only be discriminated by the third dimension (not shown).

Discussion et perspectives

La bactérie *Escherichia coli* est maintenant étudiée depuis plus de 100 ans ; elle constitue un organisme modèle et la quantité d'informations à notre disposition est très importante. Ces connaissances nous permettent de mieux comprendre comment la bactérie s'adapte à son environnement et réagit face aux différents stress qu'elle peut rencontrer. Cette faculté d'adaptation provient majoritairement de la régulation de l'expression de ses gènes, donc de leurs régulations. Le réseau de régulation d'*E. coli* est probablement celui qui est le mieux connu, et recense 4405 ORFs (Open Reading Frame ; Blattner *et al.*, 1997). De plus, 8% de ces gènes, soit environ 300 gènes sont des facteurs de transcription (Pérez-Rueda & Collado-Vides, 2000). Parmi ces régulateurs, certains sont appelés globaux car ils modulent l'expression de plusieurs centaines de gènes ; c'est par exemple le cas de CRP qui interagit avec environ 200 promoteurs (Grainger & Busby, 2008). Ainsi, ils sont des nœuds essentiels du réseau de régulation génique. Au cours de ma thèse, je me suis donc intéressé à la régulation d'un de ces régulateurs globaux, mais aussi aux conséquences de la disponibilité de la machinerie globale d'expression. En effet, le niveau d'expression de cette machinerie s'adapte aux conditions environnementales, notamment au taux de croissance.

Bien que son régulon soit important et beaucoup étudié (Lacour & Landini, 2004 ; Patten *et al.*, 2004 ; Weber *et al.*, 2005 ; Dong *et al.*, 2008 ; Dong & Schellhorn, 2009), peu d'informations sont connues concernant le contrôle de l'expression du régulateur global Crl. Nous avons tout d'abord montré que le gène *crl* présente une expression transitoire lors de la phase exponentielle de croissance. Cette expression précède l'accumulation du facteur σ^S , dont il module l'activité. En effet, lorsque Crl se fixe à σ^S il augmente son activité transcriptionnelle et favorise son recrutement par le coeur de l'ARN polymérase (Pratt & Silhavy, 1998 ; England *et al.*, 2008). Il a été montré que Crl a un effet plus important sur RpoS lorsque celui-ci est en faible concentration (Robbe-Saule *et al.*, 2007). Ces résultats suggèrent que l'expression transitoirement forte de *crl* lors de la phase exponentielle favoriserait la transcription par σ^S alors que ce dernier est en faible concentration.

Nous avons ensuite mis en évidence deux régulations du gène *crl* : une activation par la nucléoprotéine Fis et une répression par le régulateur catabolique CRP. Ainsi, l'expression de *crl* est diminuée dans un mutant *fis*. Les résultats *in vitro* de S. Lacour montrent que la nucléoprotéine Fis se fixe à plusieurs endroits sur la région promotrice de *crl* et active

directement la transcription de *crl*. Or la protéine Fis s'accumule de façon importante lors de la phase exponentielle (Azam *et al.*, 1999), ce qui correspond au moment où l'expression de *crl* est maximale. Il a été montré récemment que Fis active la transcription d'un gène RpoS-dépendant (Typas *et al.*, 2007). Le fait qu'il stimule également l'expression de *crl* renforce son rôle dans la préparation à la phase stationnaire.

Le complexe CRP-cAMP inhibe la transcription de *crl* sans se fixer sur son promoteur. Ceci suggère une régulation indirecte de *crl*. CRP est un des régulateurs majeur chez *E. coli*, il régule directement près de 200 gènes (Martínez-Antonio & Collado-Vides, 2003; Grainger *et al.*, 2005). Parmi eux, se trouve une cinquantaine de facteurs de transcription, il est donc probable que CRP régule aussi l'expression de *crl* même s'il ne se lie pas à son promoteur, en contrôlant la transcription d'un de ces facteurs. De plus, il a été montré que RpoS, le partenaire privilégié de Crl, était lui aussi réprimé par CRP (Lange & Hengge-Aronis, 1994).

Hormis ces régulations spécifiques, *crl* semble également être affecté par l'environnement. En effet, son pic d'expression a toujours lieu à la même absorbance de la culture et peut être répété en effectuant des dilutions successives lors de la croissance. Une des explications possibles serait qu'une molécule signal s'accumule dans le milieu et réprime *crl* à partir d'une concentration critique. Nous avons montré que les systèmes à deux composantes ne sont pas impliqués dans la détection d'une telle molécule et n'affectent pas la transcription de *crl*. Des résultats antérieurs effectués du laboratoire suggèrent un rôle de l'indole et du Quorum Sensing sur l'expression de *crl* (Lelong *et al.*, 2007). Néanmoins dans nos conditions, aucun des mutants impliqués dans le quorum sensing n'a montré d'effet sur l'expression de *crl*; seul l'ajout d'indole et de tryptophane (un de ses précurseurs) affecte légèrement son expression.

Une autre hypothèse concerne le rôle de la machinerie globale sur l'expression génique. En effet, la plupart des régulateurs globaux montrent un profil de transcription similaire à celui de *crl*, avec une induction lors de la phase exponentielle. Nous savons également que la composition cellulaire varie en fonction du taux de croissance. Ces pics d'expression apparaissent lors d'une reprise de croissance, il est donc possible que des changements d'activité ou de concentration de la machinerie globale soient à l'origine de ce type de profil. Grâce à un modèle mathématique développé au laboratoire, il a été montré que le

niveau d'expression de la machinerie globale d'expression, en particulier la concentration des ribosomes et de l'ARN polymérase peut expliquer de tel profil.

Ces observations sur l'impact de la machinerie globale ont fortement influencé mes travaux de thèse. C'est pourquoi je me suis intéressé à contrôler son activité et à caractériser les conséquences de tels changements sur la croissance et l'expression génique chez *E. coli*. Un des objectifs étant de vérifier si le taux de croissance est conditionné par la disponibilité de la machinerie d'expression génique.

Notre première approche consistait à moduler l'activité traductionnelle d'*E. coli*, en agissant sur la dimérisation des ribosomes et leur inactivation. Nous avons ainsi mesuré l'activité traductionnelle dans différents mutants des facteurs de dimérisation RMF, HPF et YfiA. Nos résultats se sont montrés peu encourageants, la traduction d'un gène rapporteur n'étant pas fortement affectée et la croissance absolument pas modifiée. Le rôle de ses protéines ayant lieu plutôt pendant la phase stationnaire, nous avons surexprimé la protéine RMF dans le but d'inactiver les ribosomes en phase exponentielle. Cette surexpression n'affecte que partiellement l'activité de notre système rapporteur et aucunement la croissance. Ces résultats montrent qu'il est difficile de moduler la dimérisation des ribosomes. Bien que les principaux acteurs de cette dimérisation aient été identifiés, notamment grâce aux travaux de A. Wada et de son équipe, ce n'est pas le cas pour leur dynamique d'action et d'expression. Néanmoins pour poursuivre cette étude, il pourrait être intéressant de construire une souche dans laquelle chacun de ces facteurs est absent, pour par la suite pouvoir les surexprimer indépendamment ou conjointement.

Notre contrôle de la traduction n'ayant pas donné les résultats escomptés, nous nous sommes intéressés à l'autre acteur majeur de la machinerie globale, l'ARN polymérase. *E. coli* possède une seule ARN polymérase, dont les sous-unité limitantes sont $\beta\beta'$. De plus, elles sont codées par un seul opéron *rpoBC*. Nous avons donc placé cet opéron sous le contrôle d'un promoteur inductible par l'IPTG, et ainsi pu contrôler le niveau d'ARN polymérase intracellulaire. Notre premier résultat démontre que la concentration d'ARN polymérase influence le taux de croissance d'*E. coli* de façon dose dépendante. En effet, moins il y a d'ARN polymérase exprimée moins les cellules se divisent rapidement. Néanmoins, une forte surexpression des gènes codant ces sous-unités de l'ARN polymérase ne permet pas d'obtenir un taux de croissance supérieur à une souche sauvage et n'affecte

pas la morphologie cellulaire. A l'inverse, lorsque la concentration en ARN polymérase est très faible, les cellules arrêtent de se diviser pour former de longs filaments. Les résultats du séquençage des ARN messagers, à différents niveaux d'ARN polymérase, montrent que la plupart des classes fonctionnelles sont affectées par la variation de la quantité d'ARN polymérase. Néanmoins, pour cette expérience de RNA-seq nous n'avons pas effectué de réplicats, c'est pourquoi nous ne pouvons étudier quantitativement l'effet de l'ARN polymérase sur chaque gène. Il serait intéressant d'étudier plus finement l'expression des gènes responsables de la division cellulaire comme les gènes *fts* ou de la réponse SOS, dont la perturbation pourrait expliquer la filamentation des bactéries.

Jusqu'à présent il était possible de contrôler le taux de croissance d'*Escherichia coli* en modifiant des paramètres environnementaux, par exemple la qualité nutritive du milieu, ou en ajoutant des molécules toxiques comme des antibiotiques. Avec la souche que nous avons développée au cours de ma thèse, il est possible de fixer un taux de croissance indépendamment de l'environnement. Ceci ouvre de nombreuses perspectives, qu'elles soient du domaine fondamental ou dans des domaines plus appliqués, comme les bioproductions. Nous avons vu ainsi que les 7 facteurs sigma étaient en compétition les uns avec les autres pour s'associer au cœur de l'ARN polymérase. Notre souche permet d'étudier cette compétition *in vivo* en étant désormais capable de modifier la concentration d'enzyme cœur. La production de protéines recombinantes ou de molécules d'intérêt est un des enjeux majeurs de la biotechnologie. Un des problèmes rencontrés est de pouvoir isoler facilement le produit qui nous intéresse du reste de la cellule. Une des méthodes utilisées consiste à faire exporter par la cellule ce produit d'intérêt dans le milieu extérieur. Néanmoins, il est nécessaire de cultiver de grandes quantités de cellules pour obtenir des rendements importants. Il est envisageable d'utiliser un système comme le nôtre, pour contrôler la croissance tout en continuant d'exprimer fortement des molécules d'intérêt. Nous avons montré que lorsque nos cellules arrêtent de croître, elles sont toujours vivantes et capables de répliquer leur chromosome. Il faudrait maintenant caractériser l'activité métabolique de notre souche afin de voir si les cellules sont toujours métaboliquement actives, lorsqu'elles cessent de se diviser. Si c'est le cas, notre souche pourrait être utilisée comme une usine de production biologique, utilisant les ressources du milieu uniquement pour le métabolisme et non plus pour produire de la biomasse.

Annexe

Publication n°4 :

**A genome-wide screen for identifying all regulators of
a target gene**

A genome-wide screen for identifying all regulators of a target gene

Guillaume Baptist¹, Caroline Ranquet¹, Jérôme Izard¹, Corinne Pinel¹, Delphine Ropers², Hidde de Jong², Johannes Geiselmann^{1*}

¹ Université Joseph Fourier, Laboratoire Adaptation et Pathogénie des Microorganismes, La Tronche, 38700, France

² INRIA Grenoble–Rhône-Alpes, Saint Ismier, 38334, France

* Correspondence : Johannes Geiselmann, Institut Jean Roget, LAPM, UMR5163, Campus Santé, Université Joseph Fourier, Domaine de la Merci, 38700 La Tronche, France, Tel: (+33) 4 76 63 74 79; Fax: (+33) 4 76 63 74 97; Email: hans.geiselmann@ujf-grenoble.fr

ABSTRACT

We have developed a new screening methodology for identifying all genes that control the expression of a target gene through genetic or metabolic interactions. The screen combines mutant libraries with luciferase reporter constructs. Instead of a static picture of gene expression, this method allows dynamical monitoring in different environmental conditions. Mutants with interesting phenotypes can thus be selected based on multiple criteria, and the expression dynamics of the target gene can be extensively characterized. We apply the method to identify the genes that control the expression of the *acs* gene of *Escherichia coli*. We confirm most of the known genetic regulators, including CRP-cAMP, IHF, and components of the phosphotransferase system (PTS). In addition, we identify new regulatory influences, many of which involve metabolic intermediates or metabolic sensing, such as the genes *pgi*, *pfkA*, *sucB*, and *lpdA* encoding enzymes in glycolysis and the TCA cycle. The method is applicable to any promoter fused to a luminescent reporter gene in combination with a deletion mutant library.

INTRODUCTION

The adaptation of bacteria to changes in their environment is controlled on the molecular level by a large and complex regulatory network involving genes, mRNAs, proteins and metabolites (1). Understanding, and therefore eventually predicting, the dynamics of this regulatory network is a general goal of biology, and in particular of the field of systems biology. The basic ingredient of this type of research is knowing the topology of the underlying regulatory network. An increasing number of such regulatory connections are documented in bioinformatics databases such as EcoCyc (<http://EcoCyc.org>), which lists more than 5345 regulatory interactions for *E. coli* (2).

Different methods are used to identify and characterize these interactions. Evidence for specific interactions can come from *in vitro* measurement of physical DNA-protein interactions (footprinting, gel retardation assay). Computational tools can predict interactions based on consensus sequence. Transcriptomics (DNA microarrays) is a popular approach to study transcription on a global scale (3). Typically, this technique allows detecting the targets of a particular regulator by studying the effect of the deletion of this regulator on the expression of all genes of a genome. ChIP-Chip is another global approach, identifying direct interactions of a specific transcription factor through the detection of

chromosome-wide DNA binding (4). All these tools aim at identifying regulons, i.e., all the targets of a given regulator (5). However, to our knowledge, no global method exists to determine the regulators that control, directly or indirectly, the expression of a particular target gene. We present such a method in this manuscript.

We use the GFP and luciferase reporters to measure promoter activities in living cells (6,7). These constructs provide not only a static picture of expression levels, but allow measuring the *dynamics* of gene expression with a high temporal resolution (8). The luciferase system is particularly useful since it allows measuring gene expression in a colony growing on solid medium. We have developed a technique for efficiently transforming 3571 *E. coli* single-gene knockout mutants (9) with the appropriate reporter plasmid. The screening strategy consists in comparing the luminescence of mutant colonies with the luminescence of a wild-type colony: a significant difference suggests that the deleted gene controls, directly or indirectly, the expression of the target gene.

We have applied this method to identify regulators that control the expression of the *acs* gene, coding for the acetyl coenzyme A synthetase (10). This enzyme converts acetate to acetyl coenzyme A. We have chosen this gene for three reasons. First, this enzyme plays a key role in the “acetate switch”: when bacteria grow rapidly on glucose, they excrete acetate which is subsequently used as a carbon source when glucose is exhausted (11). Acetyl coenzyme A synthetase is thus at the center of an inversion of metabolic flux from glycolysis to neoglucogenesis. Second, the accumulation of acetate in the culture medium is an important problem in industrial fermentations since this organic acid inhibits cell growth and recombinant protein production. Different approaches are currently developed to reduce acetate accumulation by modifying central metabolic pathways (12). Lastly, the *acs* gene was also selected because its complex regulation has been extensively described and thus allows a thorough validation of the method (13-15). Briefly, transcription of this gene is repressed by the nucleoproteins Fis and IHF and positively regulated by the CRP-cAMP complex. The latter activation implies the well-studied carbon catabolite repression mechanism where the glucose concentration controls the formation of the CRP-cAMP complex by adjusting the cAMP production via the phosphotransferase system (PTS) (16). The *acs* gene is thus at the end of a well-known, complex regulatory cascade, an ideal property for our purpose.

Our screen confirms known regulatory mechanisms of *acs* expression. In particular, we confirm the necessity of a functional CRP-cAMP complex for the expression of *acs*, as well as the modulation of *acs* promoter activity by IHF. The screen also highlights that glucose repression of the *acs* promoter cannot be completely explained by cAMP-mediated mechanisms and inducer exclusion, a phenomenon already observed for other *E. coli* promoters (17). We also find new regulators that establish a tight links between metabolism and gene expression, most notably genes encoding enzymes in glycolysis and the TCA cycle, such as *pgi*, *pfkA*, *sucB*, and *lpdA*. We confirmed some of these novel regulators by means of independent quantitative RT-PCR experiments.

MATERIAL AND METHODS

Strains

The screen uses the Keio collection, comprising 3985 single gene deletions of the *E. coli* K-12 strain BW25113, called wild-type strain (9).

Plasmids

The *nrfA-acs* intergenic region was amplified by PCR using primers 5'-TCCTCTCGAGAGGGGCTTCATCCGAAT-3' (*XhoI* site underlined) and 5'-TTTGGGATCCGCTTTTGTTCCTTGAGG-3' (*BamHI* site underlined) and the BW25113 strain as a template. The PCR product, digested with *XhoI* and *BamHI*, was inserted into a *luxCDABE* plasmid backbone without promoter to give p-*acs-lux*. A strong ribosome binding site drives efficient expression of the reporter gene. The *bla* gene (conferring ampicillin resistance) and the *acs* promoter region of the previous plasmid p-*acs-lux* were amplified by PCR using primers -TTTGGGATCCGCTTTTGTTCCTTGAGG-3' (*BamHI* site underlined) and 5'-TCTAAAGTGAGCTCGAGTAACTTGGTCTGACAG-3' (*SacI* site underlined). The PCR product, digested with *SacI* and *BamHI*, was inserted into pUA66 (8) to give p-*acs-gfp*. The complete sequences of these plasmids are available upon request.

High-throughput transformation

All steps were performed in 96-well plates. 135 plates, 6 days and 4 persons were needed to transform 3571 mutants. All volumes below are indicated for one well. Mutants were inoculated from frozen stock in sterile microplates containing 200 μ l per well of Luria-Bertani broth (LB) with kanamycin (50 μ g/ml) and placed at 37°C under shaking (200 rpm) overnight. The next day, 4 μ l of the pre-culture were added to new plates containing 400 μ l of LB with kanamycin (50 μ g/ml) and placed at 37°C under shaking for 3 hours. We then transferred 50 μ l of these exponentially growing cells to new plates containing 50 μ l of ice cold TSS 2X (LB containing 10% PEG, 5% DMSO, and 20-50 mM Mg^{2+} at a final pH of 6.5) (18) and the mixture was kept on ice for 45 min before adding 4 μ l of plasmid DNA (50 ng/ml). The plates were left on ice for another 10 minutes before applying a heat shock (90 seconds at 42°C) and returned on ice for an additional 10 minutes of incubation. We finally added 1 ml of LB containing ampicillin (100 μ g/ml) and the plates were placed at 37°C under shaking overnight. The following day, we spotted the transformed mutants onto two different petri dishes containing either LB-agar-ampicillin (100 μ g/ml) or minimal medium M9-agar-0.3% glucose (p/v)-ampicillin (100 μ g/ml). Petri dishes were incubated at 37°C overnight and the luminescence of the colonies was measured the following days using an intensified camera (Photonic Science). The quantification of the luminescence of the 3500 colonies of the condition "LB day 1" was performed semi-automatically by means of a custom-made Matlab® scripts (Image processing toolbox). Briefly, we used functions that correct side effects and compute an appropriate threshold on the images allowing the detection of the colonies. Then, we computed the median of luminescence pixels belonging to a specific colony.

Measuring the dynamics of gene expression

The dynamics of gene expression was measured for cells growing for 10 hours in a 96-well microplate. The microplate reader (Perkin Elmer, Fusion alpha FP-HT) allows growth with shaking and temperature control (37°C). We measured the cell density (OD at 600 nm) every 30 min and luminescence or fluorescence (480 nm/520 nm) every 3 minutes. Unless otherwise specified, the wells in the microplate contain 180 µl of minimal medium (M9) supplemented with 0.03% glucose (p/v), 0.03% sodium acetate (p/v), and ampicillin (100 µg/ml). When indicated, we added cyclic AMP (5 mM) at the beginning of the experiment. The wells were inoculated with 20 µl of stationary phase cultures (10x dilution) grown in LB-ampicillin (100 µg/ml). The fluorescence background of the bacterial cells was assessed by measuring the fluorescence of mutant bacteria transformed with a promoterless reporter plasmid.

Quantification of the expression profiles

In order to compare the strength of *acs* induction at the entry into stationary phase as measured by the luminescence and fluorescence reporter systems, we have calculated the area under the expression profile for four hours following entry into stationary phase (see Figure 4B for a schematic representation). This quantification was carried out using custom-made Matlab® scripts that perform the following tasks: (i) Remove outliers from the data. (ii) Determine the time of entry into stationary phase by classifying the expression profiles into different classes and fitting segment splines to the signal (fluorescence or luminescence) and the absorbance measurements. The entry into stationary phase is, e.g., indicated by a break in the absorbance curve. (iii) Correct the fluorescence expression profiles for the autofluorescence of the cells by subtracting corresponding profiles of control strains. No background correction is necessary for luminescence data. (iv) Numerically integrate the expression curve for 240 min, starting at the entry into stationary phase. The expression value was set to zero at the entry into stationary phase and the integral thus gives a measure of the induction strength of the *acs* promoter in the particular mutant strain. (v) Normalize the integral by the absorbance of the sample at entry into stationary phase.

qRT-PCR

The strains were grown in a microplate containing M9 0.03% glucose and 0.03% acetate (preculture LB). The cells ($\sim 10^8$) were harvested either before glucose exhaustion (exponential growth on glucose) or after glucose exhaustion (growth on acetate). Total mRNA was protected using the RNeasy Protect Bacteria Reagent (Qiagen) and then isolated using the RNeasy mini kit (Qiagen) according to the recommendations of the manufacturer. Potential trace quantities of DNA were removed using the turbo DNase (Ambion). Samples were protected from RNase using RNaseOut ribonuclease inhibitor (Invitrogen). 1 µg of RNA was reverse transcribed using SuperScript™ II Reverse Transcriptase (Invitrogen) according to the protocol of the manufacturer. Briefly, a 25 µl reaction mixture was incubated in a T3000 thermocycler (Biometra) for 10 min at 25°C, 50 min at 42°C and 15 min at 70°C. The primers (Eurogentec) used in the present study were as follows: *acs*: 5'-ACAGTTCTGGTGACGGTTCC-3' and 5'-ACAGTTCTGGTGACGGTTCC-3'. We have used the

rrsD gene as internal control for relative quantification: *rrsD*: 5'-GCTACAATGGCGCATACAAA-3' and 5'-TTCATGGAGTCGAGTTGCAG-3'. Quantitative PCR was performed in a StepOnePlus Real-Time PCR System (Applied Biosystems) using MESA Green qPCR Master Mix (Eurogentec) according to the instructions of the manufacturer. Briefly, 25 µl reactions mixtures were incubated for 10 min at 95°C and 40 PCR cycles (15 s at 95°C and 1 min at 60°C). PCRs were run in triplicate. Raw data were transformed into threshold cycle (Ct) values. The relative strength of *acs* expression for each mutant, compared to wild-type, was calculated by the comparative Ct Method ($\Delta\Delta C_t$).

RESULTS

A screening strategy for identifying the regulators of *acs*

Our screening strategy is outlined in Fig. 1. To measure *acs* promoter activity we have constructed a reporter plasmid containing a transcriptional fusion of the *acs* promoter region (394 bp) to the *luxCDABE* operon. This operon codes for the heterodimeric luciferase protein (*luxAB*) and the enzymes (*luxCDE*) necessary for producing the luciferase substrate, a long-chain aldehyde (19). We have chosen this reporter system because of the highly sensitive luminescence signal, the absence of background, the continuous production of bioluminescence without added substrate, and the possibility to measure luminescence at the colony level. The reporter plasmid was transformed into the 3911 single-gene knockouts of *Escherichia coli* (Keio collection) using our high-throughput transformation procedure (see Methods). Transformation efficiency was 91%: we have thus obtained 3571 transformants. We spotted the transformed bacteria on solid media and acquired the luminescence images of the colonies that appeared on the plates.

The luminescence of the colonies changes as a function of culture time and growth condition. The powerful features of our screen are that we can monitor the dynamics of gene expression at the colony level (see an example in Supplementary Information 2: Movie) and adapt growth conditions to be most relevant to the gene under investigation. The *acs* promoter is activated after exhaustion of the preferred carbon source. We therefore measured the luminosity of colonies after 24h and 48h of growth on two different media: rich LB medium and minimal glucose medium. We thus obtained four complete sets of luminosity images of the plates. We quantified the luminosities of the 3571 colonies after 24h of growth on LB plates (Supplementary Data). Fig. 1 shows the corresponding distribution of luminosities. Automatic quantification was not possible for the other growth conditions, but the complete dataset is available in Supplementary Data. From these four batches of luminescence pictures, we have selected 517 (15%) colonies for a detailed kinetic analysis. We also include data about the growth phenotype: among the 3571 transformants, eighty auxotrophic mutants did not grow on minimal medium-agar-glucose (Supplementary Information 4).

The dynamics of *acs* expression

For the 517 pre-selected genes, we measured the dynamics of *acs* expression in liquid cultures grown in a minimal medium containing glucose and acetate in order to focus on the physiologically

important transition between those two carbons sources. The expression patterns were measured in an automated microplate reader and analyzed as previously described (20). Fig. 2A shows the expression profile of *acs* in the wild-type strain. We observe a sharp increase of *acs* transcription when glucose is exhausted. This expression pattern is very robust and independent of the reporter system since we observe an identical pattern with the *gfp* reporter carried on a plasmid with a completely different replication mechanism (Fig. 2B). Note that luminescence is preferable for screening because bacteria have a significant auto-fluorescence background, probably due to flavin mononucleotide or riboflavin (21).

Known regulators

An important criterion for the validity of a screen is that known interactions should be confirmed. The *acs* gene is positively regulated by the CRP-cAMP complex (*crp* gene) and thus by the well-studied carbon catabolite repression mechanism. The influx of glucose through the phosphotransferase system (PTS) controls adenylate cyclase (*cyaA* gene) activity and thus cAMP production. The PTS is composed of enzymes EI (*ptsI* gene), HPR (*ptsH* gene), EIIA_{glc} (*crr* gene) and EIIB/C_{glc} (the glucose transporter encoded by the *ptsG* gene). When glucose is exhausted, the four PTS enzymes are phosphorylated and EIIA_{glc} binds to and activates adenylate cyclase. The cAMP concentration quickly rises and the Crp-cAMP complex activates target promoters such as the *acs* promoter (16) (Fig. 5).

The analysis of the genes involved in glucose repression confirms that our screen reliably identifies dramatic, but also subtle changes in the regulation of the target gene. In *cyaA* and *crp* strains, there is no transcription of *acs* at all, demonstrating the strict dependence of *acs* transcription on the CRP-cAMP complex. The activity of the *acs* promoter can be restored in the *cyaA* strain by adding cAMP to the medium but, interestingly, only after glucose exhaustion (Fig. 2C). A similar phenomenon was observed before for other *E. coli* promoters (17). The *crr* and *ptsI* strains show the expected phenotype: a very low *acs* expression, although significantly higher than in the *cyaA* and *crp* strains. cAMP added to the medium clearly restores *acs* expression in these two strains confirming that these mutants cannot activate adenylate cyclase. As in the *cyaA* strain, cAMP fails to complement the *crr* strain while glucose is still present in the growth medium. In contrast, the *ptsI* strain is directly complemented by cAMP even in presence of glucose (Fig. 2D). This difference of reactivity is not explained by the current model of catabolite repression: adding cAMP to the medium should bypass the glucose repression. Consequently, the *cyaA* and *crr* mutants should behave as the *ptsI* strain, with a direct activation of *acs* even in presence of glucose.

Expression of *acs* is reported to be negatively regulated by two histone-like proteins Fis and IHF (22). The Fis protein is abundant in exponential phase (23). Since *acs* expression is already very low in this growth phase, the observed effect of the *fis* deletion is rather subtle: the *acs* expression profile is rather similar to that of the wild-type strain (data not shown). IHF acts mainly in stationary phase (11) and we accordingly observe a higher *acs* expression in late stationary phase (Fig. 2F). The perfect superposition of *acs* expression profiles observed in the *ihfA* and *ihfB* strains (each coding for a subunit of IHF and transcribed from different chromosomal locations) confirms the high accuracy and internal consistency of the method.

Metabolic influences on the reporter system

For the purpose of our screen, we want luciferase to report only the changes in gene expression of the target gene. However, the emission of light produced by the colonies does not only depend on *luxCDABE* transcription, but is also influenced by the metabolic state of the cell, such as concentrations of oxygen, FMNH₂, ATP, and NADPH. These metabolites are involved in the enzymatic reactions that lead to light emission (24). The *nuo* operon (13 genes) is a striking example of these metabolic effects on luciferase activity: all corresponding colonies show a higher luminescence on LB plates (Fig. 3A). The *nuo* operon codes for the NADH dehydrogenase involved in the respiratory chain. In the absence of a functional enzyme, reducing power accumulates (25), thus stimulating the luciferase reaction.

A systematic way to detect and eliminate the above false positives, due to the reporter system, is to validate the results with a different reporter system that is independent of the metabolic state of the cell. We therefore transformed the 517 selected mutants with the *gfp* reporter plasmid, also carrying a different origin of replication (thus controlling for mutants that would affect the plasmid copy number). We acquired the expression profile of *acs* for all selected mutants using the two reporter systems. Fig. 4A shows, for each reporter, the strength of *acs* induction, computed from the area under the induction curve. Mutants for which *acs* activation is very different in the two reporter systems (points far off the diagonal) likely represent artifacts of at least one of the reporter systems. We notably report 14 false positives along the y-axis of the scatterplot, corresponding to mutants with *acs* expression in a *gfp* reporter plasmid and no induction in a *lux* reporter, due to metabolic effects affecting light production rather than *acs* expression. We also excluded 9 mutants along the x-axis, which for unknown reasons show no *acs* expression in a *gfp* reporter plasmid.

The mutants having very low or very high *acs* induction in both reporter plasmids can be considered the true positives of the screen (Fig. 4A). As a validation of the results, we measured the transcription of several genes with an independent gene expression assay, using quantitative RT-PCR. Mutants in genes coding for the PTS (*ptsG*, *csr*, *ptsI*), cAMP production (*cyaA*), regulatory proteins (*ihfB*) and metabolic enzymes (*lpdA*, *sucB*) precisely reproduce the effects observed with the reporter genes (Fig. 4B).

Metabolic regulation of *acs* expression

An interesting observation is that many of the regulators of *acs* expression returned by the screen are genes encoding enzymes in carbon metabolism. This reminds that gene expression in bacteria is tightly interlinked with the metabolic state of the cell.

A first example of this connection revealed by the screen are mutants affecting glycolytic fluxes. According to the classical model of carbon catabolite repression, a diminution of glycolytic fluxes increases the phosphorylation of the PTS leading to an increase of the concentration of cAMP and ultimately the expression of *acs* (26). The *ptsG* mutant (which lacks the glucose transporter) follows the classical model: colonies grow more slowly on glucose in the *ptsG* strain than in the wild-type strain, and the dynamical analysis shows a higher *acs* expression level in exponential phase (Fig.

2E). Interestingly, our screen reveals that the deletion of down-stream glycolytic enzymes leads to a different and counter-intuitive expression phenotype. For example, while *pgi* and *pfkA* mutants grow more slowly on glucose, *acs* is not derepressed in exponential phase (Fig. 2E).

As a second example, several of the mutants with the lowest colony luminescence carry deletions of genes involved in the energy supply of the cell (Fig. 1), such as the genes coding for ATP synthetase (*atpC*, *atpD*, *atpE*, *atpF*) and the genes coding for enzymes of the TCA cycle: the oxoglutarate dehydrogenase complex (*sucA*, *sucB* and *lpdA*) and, to a lesser extent, succinyl-CoA synthetase (*sucCD*) and succinate dehydrogenase (*sdhCDAB*). We found that these mutants fail to grow on acetate (Supplementary Information 4). In these mutants, the luminescence profile shows an almost total absence of *acs* expression (Fig. 3B). While it might be argued that this is due to a failure in energy supply, recall from the previous section that the effect is reproduced with the fluorescence reporter and qRT-PCR (Fig. 4), thus strongly suggesting the existence of a genuine regulation.

DISCUSSION

A new method for identifying the regulators of *acs*

The classical way to dissect a regulatory network, closely linked to classical approaches in genetics, consists in disrupting a gene of interest and observing the effects on the organism. This approach identifies the downstream targets of the gene of interest. Here, we have developed a method to do the opposite: given a gene of interest, we can identify all factors, genetic and metabolic, that affect the expression of this gene. Our method relies on a high-throughput transformation of a luciferase reporter plasmid into an extensive mutant collection. Even though we have focused on transcriptional regulation, the method can certainly be extended to include translational fusions. We use the Keio-mutant collection, but the scope of the screen could be extended to include, e.g., mutants of non-coding RNAs and intergenic regions that may contain unknown ORFs.

We have chosen luciferase as a reporter because it allows easy screening at the colony level without any significant background signal. Furthermore, the luciferase signal of a colony is highly reproducible, with a standard deviation of about 5% for independent colonies of the same mutant. The enzymatic reaction of luciferase depends on cellular factors such as ATP concentration or redox potential. For a given medium, deletions that seriously affect the concentration of these factors may generate false positives. However, these can be easily excluded by using an alternative reporter system or verification by qRT-PCR.

An important novel feature of our screen is the possibility to directly measure the dynamics of gene expression during the screen or by high-throughput measurements of selected mutants in an automated plate reader. Furthermore, the original screen, as well as these additional analyses, can be carried out in many different environmental and growth conditions.

The regulation of *acs* transcription

The effects of some interesting mutants identified by the screen are summarized in Fig. 5. We have confirmed known interactions that regulate the transcription of *acs* and we have discovered interesting new directions to explore. The *acs* promoter activity is controlled by IHF and, most importantly, the CRP-cAMP complex: transcription is completely abolished without CRP or cAMP. We also confirm the involvement of the phosphotransferase system, which sets the intracellular concentration of cAMP: the *crr* and *ptsI* mutants have a very limited *acs* expression except when cAMP is added to the medium. Interestingly, even though part of the same glucose uptake machinery, we found that the *ptsG* mutant behaves differently from the other PTS mutants: it does not require the addition of exogenous cAMP to activate the *acs* expression.

We failed to induce *acs* expression in *cyaA* or wild-type strains by adding cAMP while glucose was still present in the growth medium. Those results could possibly be explained by another well-known mechanism of glucose repression: inducer exclusion. In this control system, the dephosphorylated form of EIIA_{glc}, the dominant form in the presence of glucose, binds and inhibits different sugar transport systems, thus preventing the entry of the inducing sugar (reviewed by (27)) into the cell. However, inducer exclusion does not account for the glucose repression of the *acs* transcription in the presence of exogenous cAMP. To date, no inducer of *acs* expression has been described and acetate can enter the cell by passive diffusion (28). Furthermore, the central component of inducer exclusion is EIIA_{glc}, encoded by the *crr* gene. Yet, we still observe glucose repression in a *crr* mutant in the presence of exogenous cAMP (Fig. 2D). This experiment demonstrates that neither variations of cAMP concentration nor variations of the phosphorylated state of EIIA_{glc} are sufficient to explain the glucose repression that we observe. Consequently, while confirming the involvement of the CRP-cAMP complex and the PTS in the transcription of the *acs* gene, our results show that the current two main models of carbon catabolite repression are incomplete (17). Despite more than four decades of research, our understanding of the precise mechanism that underlie the glucose effect (and more generally the carbon catabolite repression) is still far from complete and remains the subjects of debate (29-32).

Metabolism and gene expression

An important strength of our screen is the ability to uncover connections between metabolism and gene expression. In the case of *acs*, we find several novel connections between the metabolic state of the cell and the transcription of *acs*. For instance, the enzymes involved in the growth on acetate (*suc*, *atp* mutants etc...) have a genuine impact on *acs* expression, as verified by means of independent methods. Even though such a regulation makes sense biologically, the mechanism of this signal transduction remains to be discovered. Another example involves the glycolytic flux. Reducing the glucose inflow by deleting the glucose transporter (*ptsG* gene) increases the basal *acs* expression level during growth on glucose, probably by increasing the intracellular concentration of cAMP. However, the reduction of the glycolytic flux observed during growth of the *pgi* and *pfkA* mutants does not lead to an enhancement of *acs* expression in exponential phase (Fig. 2E). This result shows that neither *acs* promoter activity nor cAMP concentration are a simple function of the glycolytic flux, which

is roughly the same for *ptsG*, *pgi*, and *pfkA* strains, but depends on the exact location where glycolysis is interrupted. We do not yet know the metabolic indicator nor the sensing mechanism that transmits the state of glycolysis to the activation of the *acs* promoter.

These different observations clearly show a strong connection between the metabolic state of the cell and *acs* transcription and confirm our prediction that the genetic regulatory network of *Escherichia coli* is densely connected and that a majority of these connections pass through metabolism (33).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online: Supplementary Data (image or excel file), Supplementary Information I-IV.

ACKNOWLEDGEMENT

We would like to thank Omayya Dudin, Stephan Lacour and Yves Markowicz for many useful discussion and comments on the manuscript. Laure Giannone helped with the analysis of the data.

FUNDING

This work was funded by the EU FP6 grant EC-MOAN; the ANR BioSys grant MetaGenoReg and the Joseph Fourier University.

REFERENCES

1. Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology*, 3, 318-356.
2. Keseler, I.M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muniz-Rascado, L., Bonavides-Martinez, C., Paley, S., Krummenacker, M., Altman, T. *et al.* (2010) EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic acids research*, 39, D583-590.
3. Brown, P.O. and Botstein, D. (1999) Exploring the new world of the genome with DNA microarrays. *Nature genetics*, 21, 33-37.
4. Grainger, D.C. and Busby, S.J. (2008) Methods for studying global patterns of DNA binding by bacterial transcription factors and RNA polymerase. *Biochemical Society transactions*, 36, 754-757.
5. Neidhardt, F.C. and Savageau, M.A. (1996), *Escherichia coli and Salmonella: Cellular and Molecular Biology*. ASM Press, Washington, DC, Vol. 1, pp. 1310-1324.
6. Van Dyk, T.K., Wei, Y., Hanafey, M.K., Dolan, M., Reeve, M.J., Rafalski, J.A., Rothman-Denes, L.B. and LaRossa, R.A. (2001) A genomic approach to gene fusion technology. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 2555-2560.
7. Hakkila, K., Maksimow, M., Karp, M. and Virta, M. (2002) Reporter genes lucFF, luxCDABE, gfp, and dsred have different characteristics in whole-cell bacterial sensors. *Analytical biochemistry*, 301, 235-242.
8. Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M.G. and Alon, U. (2006) A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nature methods*, 3, 623-628.
9. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology*, 2, 2006 0008.
10. Kumari, S., Tishel, R., Eisenbach, M. and Wolfe, A.J. (1995) Cloning, characterization, and functional expression of *acs*, the gene which encodes acetyl coenzyme A synthetase in *Escherichia coli*. *Journal of bacteriology*, 177, 2878-2886.
11. Wolfe, A.J. (2005) The acetate switch. *Microbiol Mol Biol Rev*, 69, 12-50.

12. Gosset, G. (2005) Improvement of *Escherichia coli* production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. *Microbial cell factories*, 4, 14.
13. Kumari, S., Beatty, C.M., Browning, D.F., Busby, S.J., Simel, E.J., Hovel-Miner, G. and Wolfe, A.J. (2000) Regulation of acetyl coenzyme A synthetase in *Escherichia coli*. *Journal of bacteriology*, 182, 4173-4179.
14. Sclavi, B., Beatty, C.M., Thach, D.S., Fredericks, C.E., Buckle, M. and Wolfe, A.J. (2007) The multiple roles of CRP at the complex *acs* promoter depend on activation region 2 and IHF. *Molecular microbiology*, 65, 425-440.
15. Shin, S., Chang, D.E. and Pan, J.G. (2009) Acetate consumption activity directly determines the level of acetate accumulation during *Escherichia coli* W3110 growth. *Journal of microbiology and biotechnology*, 19, 1127-1134.
16. Gorke, B. and Stulke, J. (2008) Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol*, 6, 613-624.
17. Ullmann, A. and Danchin, A. (1983) Role of cyclic AMP in bacteria. *Adv. Cyclic Nucleotide Res*, 15, 1-53.
18. Chung, C.T., Niemela, S.L. and Miller, R.H. (1989) One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences of the United States of America*, 86, 2172-2175.
19. Meighen, E.A. (1991) Molecular biology of bacterial bioluminescence. *Microbiological reviews*, 55, 123-142.
20. de Jong, H., Ranquet, C., Ropers, D., Pinel, C. and Geiselmann, J. (2010) Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria. *BMC systems biology*, 4, 55.
21. Billinton, N. and Knight, A.W. (2001) Seeing the wood through the trees: a review of techniques for distinguishing green fluorescent protein from endogenous autofluorescence. *Analytical biochemistry*, 291, 175-197.
22. Browning, D.F., Beatty, C.M., Sanstad, E.A., Gunn, K.E., Busby, S.J. and Wolfe, A.J. (2004) Modulation of CRP-dependent transcription at the *Escherichia coli* *acsP2* promoter by nucleoprotein complexes: anti-activation by the nucleoid proteins FIS and IHF. *Molecular microbiology*, 51, 241-254.
23. Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S. and Ishihama, A. (1999) Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *Journal of bacteriology*, 181, 6361-6370.
24. Koga, K., Harada, T., Shimizu, H. and Tanaka, K. (2005) Bacterial luciferase activity and the intracellular redox pool in *Escherichia coli*. *Mol Genet Genomics*, 274, 180-188.
25. Pruss, B.M., Nelms, J.M., Park, C. and Wolfe, A.J. (1994) Mutations in NADH:ubiquinone oxidoreductase of *Escherichia coli* affect growth on mixed amino acids. *Journal of bacteriology*, 176, 2143-2150.
26. Hogema, B.M., Arents, J.C., Bader, R., Eijkemans, K., Inada, T., Aiba, H. and Postma, P.W. (1998) Inducer exclusion by glucose 6-phosphate in *Escherichia coli*. *Molecular microbiology*, 28, 755-765.
27. Roseman, S. and Meadow, N.D. (1990) Signal transduction by the bacterial phosphotransferase system. Diauxie and the *crr* gene (J. Monod revisited). *The Journal of biological chemistry*, 265, 2993-2996.
28. Gimenez, R., Nunez, M.F., Badia, J., Aguilar, J. and Baldoma, L. (2003) The gene *yjcG*, cotranscribed with the gene *acs*, encodes an acetate permease in *Escherichia coli*. *Journal of bacteriology*, 185, 6448-6455.
29. Crasnier-Mednansky, M. (2008) Is there any role for cAMP-CRP in carbon catabolite repression of the *Escherichia coli* *lac* operon? *Nat Rev Microbiol*, 6, 954; author reply 954.
30. Gorke, B. and Stulke, J. (2008) Is there any role for cAMP-CRP in carbon catabolite repression of the *Escherichia coli* *lac* operon? Reply from Gorke and Stulke. *Nat Rev Microbiol*, 6, 954.
31. Inada, T., Kimata, K. and Aiba, H. (1996) Mechanism responsible for glucose-lactose diauxie in *Escherichia coli*: challenge to the cAMP model. *Genes Cells*, 1, 293-301.
32. Narang, A. (2009) cAMP does not have an important role in carbon catabolite repression of the *Escherichia coli* *lac* operon. *Nat Rev Microbiol*, 7, 250.
33. Baldazzi, V., Ropers, D., Markowicz, Y., Kahn, D., Geiselmann, J. and de Jong, H. (2010) The carbon assimilation network in *Escherichia coli* is densely connected and largely sign-determined by directions of metabolic fluxes. *PLoS Comput Biol*, 6, e1000812.

TABLE AND FIGURES LEGENDS

Figure 1. A screening strategy for identifying new regulators:

The *acs* promoter region was cloned upstream of the *luxCDABE* operon on a low-copy number plasmid. This plasmid is transformed into 3571 single knock-out mutants of *E. coli* (top). The transformed bacteria were spotted onto solid medium and the luminescence of the colonies was quantified (LB-agar day 1) (middle). The histogram (bottom) shows the distribution of luminescence emitted by all colonies. The names of the 20 mutants with the highest/lowest luciferase activity (in red on the histogram) are listed on the sides. In these mutants, light emission is significantly different from the wild-type (0.95 confidence), as determined by fitting a normal distribution to the histogram and noting that the mean of the distribution corresponds to wild-type levels. The broken vertical lines indicate two standard deviations above and below the mean. Among the mutants with the lowest luciferase activity, we find 7 genes (*atpC*, *atpD*, *atpE*, *atpG*, *sucA*, *sucB*, *sdhB*, *lpdA*) involved in energy supply and four genes involved in catabolite repression (*cyaA*, *crp*, *crr*, *ptsI*). Among the mutants with the highest luciferase activity, our screen detects a transcriptional regulator (*hha*) and 10 genes of the *nuo* operon.

Figure 2. Dynamics of gene expression of known mutants in a liquid minimal medium containing glucose and acetate:

A. The expression profile of the *acs* luciferase reporter plasmid shows that *acs* expression is induced when glucose is exhausted. B. The *acs* expression profile using GFP as a reporter is identical to the profile observed with luciferase (A). C. *acs* expression is completely absent in *crp* and *cyaA* strains. cAMP added to the medium (5mM) at the beginning of the experiment complements the *cyaA* strain but glucose repression is maintained. D. Very low expression of *acs* in the *crr* and *ptsI* strains. cAMP added to the medium at the beginning of the experiment directly complements the *ptsI* strain during growth on glucose, whereas the *crr* strain only responds once glucose is exhausted. E. In the *ptsG* mutant, *acs* induction is delayed because glucose is consumed more slowly. Moreover, *acs* expression during glucose consumption is much stronger than in the wild-type strain. The glycolysis mutants *pgi* and *pfkA* also grow more slowly, but *acs* expression remains very low in exponential phase. F. The expression profiles of the *ihfA* and *ihfB* strains show stronger *acs* expression than the wild-type during growth on acetate. The highly similar expression kinetics of the two mutants (with genes located at different sites on the chromosome) shows the high internal consistency of the method.

Figure 3. Luciferase and the energy supply:

A. Higher level of luminescence of different mutants of the *nuo* operon. This effect is probably due to the dependence of the luciferase reaction on reducing power. The *nuo* operon codes for the NADH dehydrogenase involved in the respiratory chain. In the absence of a functional enzyme, reducing power accumulates and thus stimulates the luciferase reaction. The uniform increase in luminescence

of all genes in the operon illustrates the high reproducibility of the screen. B. The *atpC* and *sucB* mutants do not induce *acs* (producing light) when glucose is exhausted.

Figure 4. Comparing luminescence, fluorescence and qRT-PCR:

A. The area under the expression curve after glucose exhaustion is a good measure of the induction of *acs* expression. The areas were computed for all measured luminescence and fluorescence profiles. The scatterplot shows the correspondence between induction measured by the luminescence and fluorescence reporters. Mutants beyond the dotted lines (four fold deviation from the correlation line) show a significant discrepancy between the two reporter systems. Mutants for which the fluorescence induction is normal, but luminescence low (red), probably affect the luminescence reporter system. Mutants that affect catabolite repression, the TCA cycle, or ATP synthetase show very low fluorescence and luminescence (green). B. Direct measurement of relative mRNA levels of selected mutants using qRT-PCR. All mRNA levels were measured 30 min after glucose exhaustion except for the first two samples, marked Wt exp and *ptsG* exp, that were measured during growth on glucose. All concentrations were normalized to the expression level of the induced wild-type strain. The *acs* mRNA is almost undetectable in the wild-type strain in exponential phase, whereas the *ptsG* mutant shows a clear signal. No *acs* expression is observed in the *cyaA* mutant (*cyaA*), but expression is restored when the medium is complemented with cAMP (*cyaA* cAMP). As expected, the *crr*, *ptsI*, *sucB* and *lpdA* strains show reduced *acs* expression in comparison with the wild-type strain, and the *ihfB* mutant has stronger *acs* expression.

Figure 5. The regulation of *acs* expression

The expression of *acs* is regulated by genes involved in glucose repression (*cyaA*, *crp*, *crr* and *ptsI*). A modification of the glycolytic flux has indirect consequences on *acs* expression (*ptsG*, *pgi*, *pfkA*). Genes of the TCA cycle are needed for acetate utilization and mutants in these enzymes shut down *acs* transcription by a yet unknown mechanism. The red and green routes are schematic representations of the flux direction when cells grow on glucose and acetate, respectively.

FIGURES

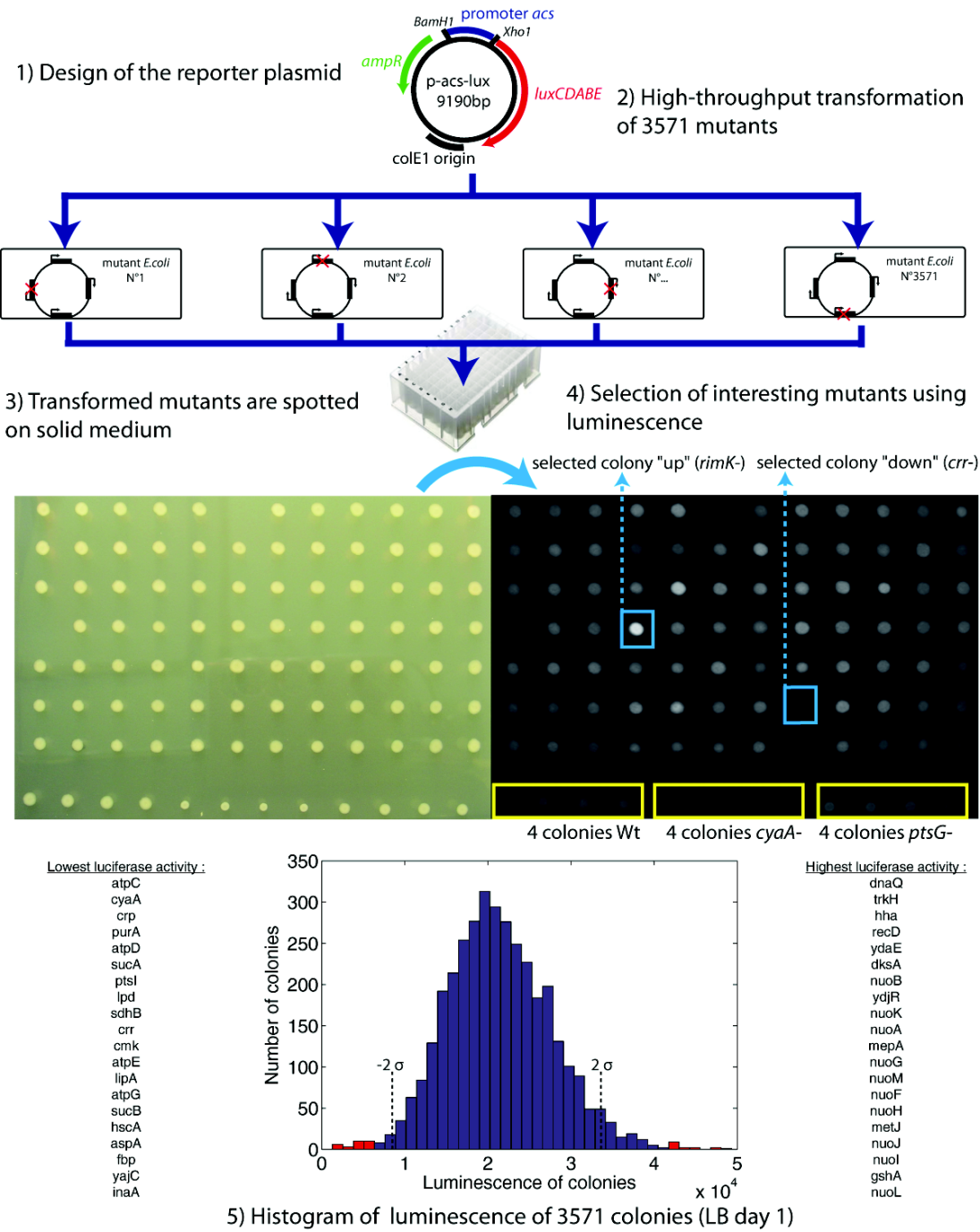


Figure 1

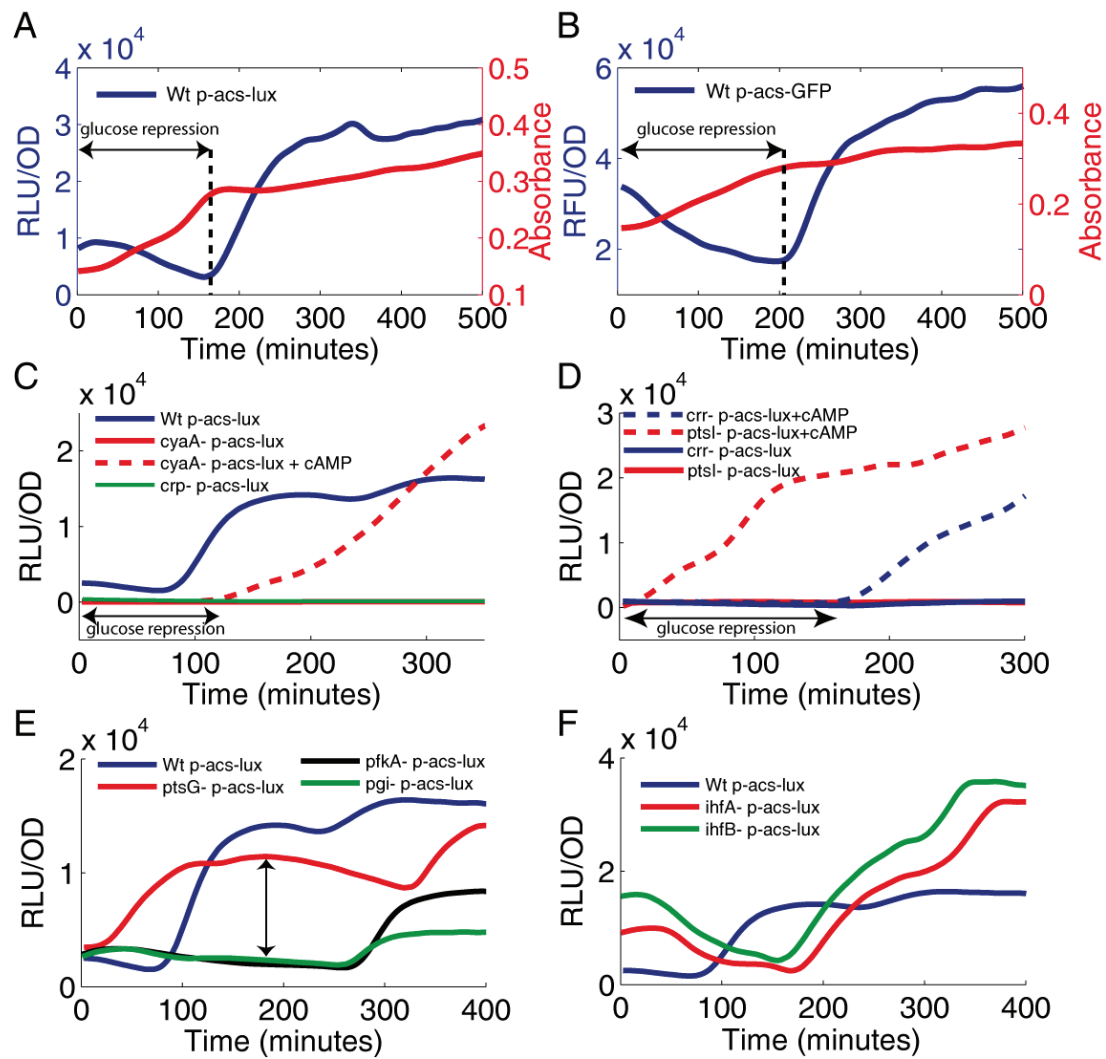


Figure 2

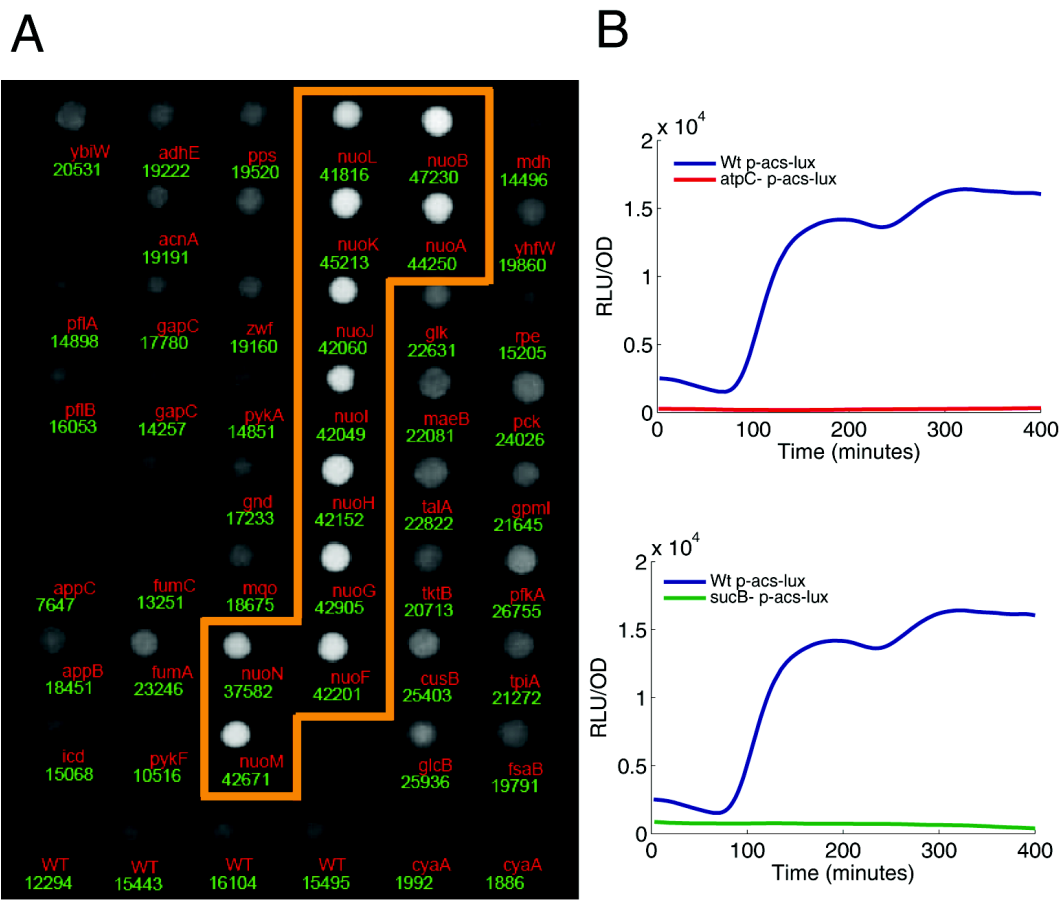


Figure 3

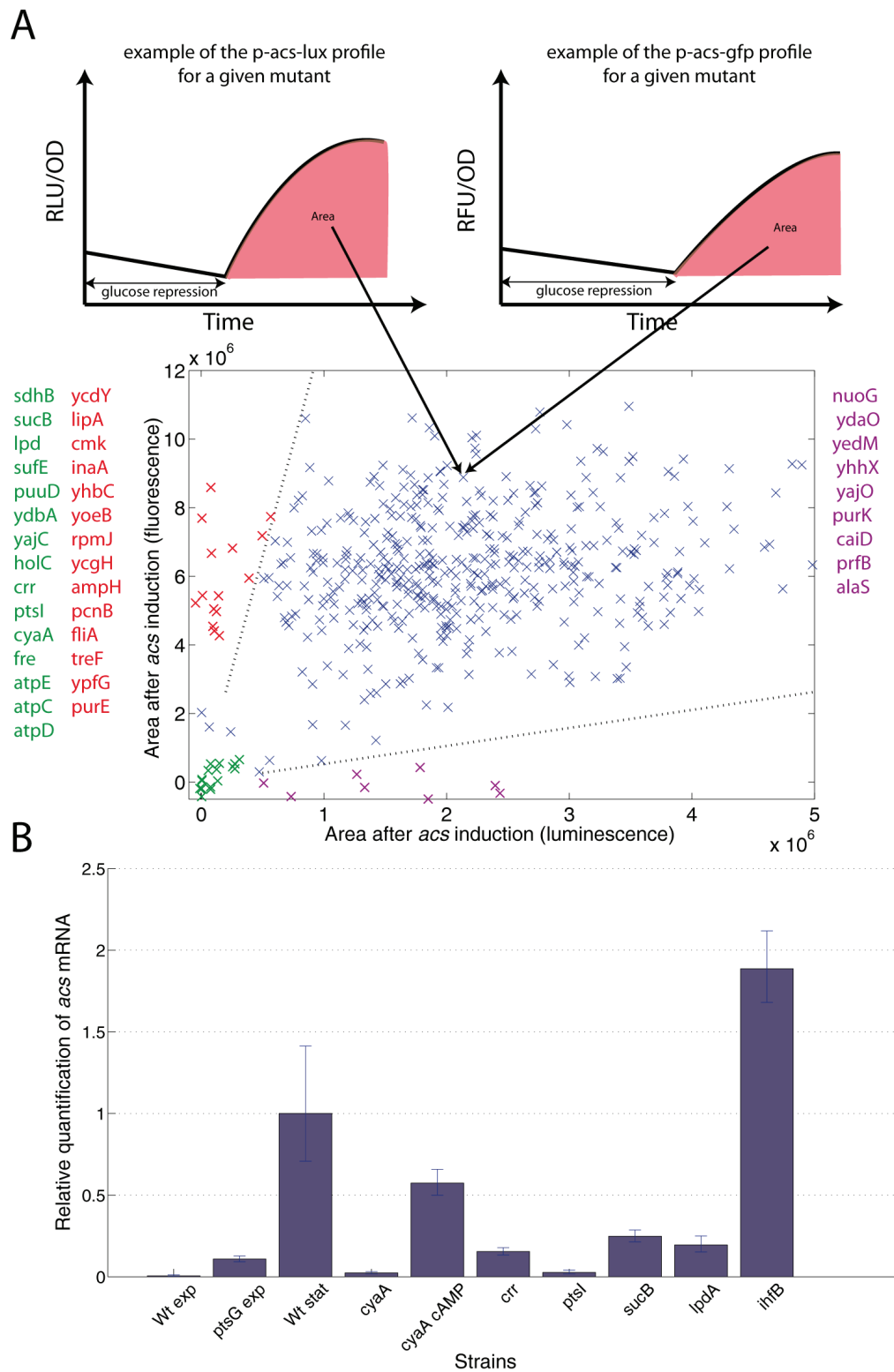
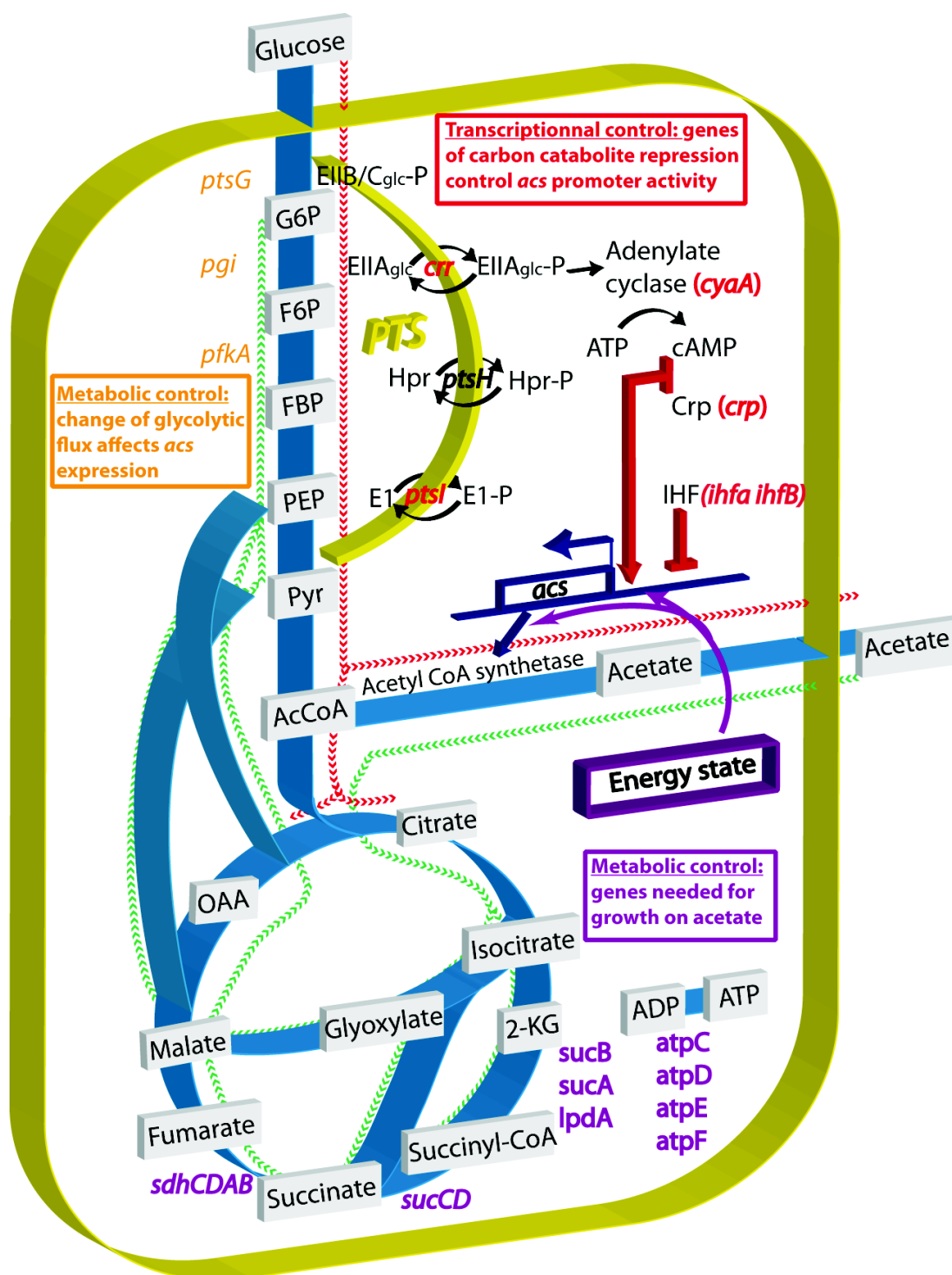


Figure 4



Bibliography

- Adhya, S. & Gottesman, M. (1982). Promoter occlusion: transcription through a promoter may inhibit its activity. *Cell*, *29*(3), 939–944.
- Agafonov, D. E., Kolb, V. A., Nazimov, I. V., & Spirin, A. S. (1999). A protein residing at the subunit interface of the bacterial ribosome. *Proc Natl Acad Sci U S A*, *96*(22), 12345–12349.
- Agafonov, D. E. & Spirin, A. S. (2004). The ribosome-associated inhibitor A reduces translation errors. *Biochem Biophys Res Commun*, *320*(2), 354–358.
- Ahmed, T., Shimizu, T. S., & Stocker, R. (2010). Microfluidics for bacterial chemotaxis. *Integr Biol (Camb)*, *2*(11-12), 604–29.
- Aiso, T., Yoshida, H., Wada, A., & Ohki, R. (2005). Modulation of mRNA stability participates in stationary-phase-specific expression of ribosome modulation factor. *J Bacteriol*, *187*(6), 1951–1958.
- Alper, H. & Stephanopoulos, G. (2007). Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab Eng*, *9*(3), 258–67.
- Arnqvist, A., Olsén, A., Pfeifer, J., Russell, D. G., & Normark, S. (1992). The Crl protein activates cryptic genes for curli formation and fibronectin binding in *Escherichia coli* HB101. *Mol Microbiol*, *6*(17), 2443–2452.
- Azam, T. A., Iwata, A., Nishimura, A., Ueda, S., & Ishihama, A. (1999). Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J Bacteriol*, *181*(20), 6361–6370.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*, *2*, 2006.0008.
- Bachmann, B. J. (1990). Linkage map of *Escherichia coli* K-12, edition 8. *Microbiol Rev*, *54*(2), 130–97.
- Baracchini, E. & Bremer, H. (1988). Stringent and growth control of rRNA synthesis in *Escherichia coli* are both mediated by ppGpp. *J Biol Chem*, *263*(6), 2597–602.

- Barry, G., Squires, C. L., & Squires, C. (1979). Control features within the *rplJL-rpoBC* transcription unit of *Escherichia coli*. *Proc Natl Acad Sci U S A*, 76(10), 4922–6.
- Bedwell, D. M. & Nomura, M. (1986). Feedback regulation of RNA polymerase subunit synthesis after the conditional overproduction of RNA polymerase in *Escherichia coli*. *Mol Gen Genet*, 204(1), 17–23.
- Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol*, 62(3), 293–300.
- Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., & Shao, Y. (1997). The Complete Genome Sequence of *Escherichia coli* K-12. *Science*, 277(5331), 1453–1462.
- Bokal, A. J., Ross, W., Gourse, R. L., & IV, A. B. (1995). The transcriptional activator protein FIS: DNA interactions and cooperative interactions with RNA polymerase at the *Escherichia coli* rrnB P1 promoter. *J Mol Biol*, 245(3), 197–207.
- Bougdour, A. & Gottesman, S. (2007). ppGpp regulation of RpoS degradation via anti-adaptor protein IraP. *Proc Natl Acad Sci U S A*, 104(31), 12896–901.
- Bougdour, A., Lelong, C., & Geiselmann, J. (2004). Crl, a low temperature-induced protein in *Escherichia coli* that binds directly to the stationary phase sigma subunit of RNA polymerase. *J Biol Chem*, 279(19), 19540–19550.
- Braatsch, S., Helmark, S. r., Kranz, H., Koebmann, B., & Jensen, P. R. (2008). *Escherichia coli* strains with promoter libraries constructed by Red/ET recombination pave the way for transcriptional fine-tuning. *Biotechniques*, 45(3), 335–337.
- Bremer, H., Baracchini, R. L., & Ryals, J. (1987). Control of RNA synthesis in bacteria. In M. F. Tuite & M. Picard (Eds.), *Genetics of Translation: New Approaches (Nato a S I Series Series H, Cell Biology)*, volume 14 (pp. 63–74). New York, N.Y.: Springer-Verlag.
- Bremer, H. & Dennis, P. P. (1996). Modulation of chemical composition and other parameters of the cell by growth rate. In F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, & H. E. Umbarger (Eds.), *Escherichia coli and Salmonella* (ASM Press ed.), volume 2 chapter 97, (pp. 1553–1569). Washington, D.C: ASM Press.
- Brewer, B. J. (1988). When polymerases collide: replication and the transcriptional organization of the *E. coli* chromosome. *Cell*, 53(5), 679–86.
- Burgess, R. R. (1969). Separation and characterization of the subunits of ribonucleic acid polymerase. *J Biol Chem*, 244(22), 6168–76.

- Cabrera, J. E. & Jin, D. J. (2003). The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. *Mol Microbiol*, 50(5), 1493–1505.
- Cabrera, J. E. & Jin, D. J. (2006). Active transcription of rRNA operons is a driving force for the distribution of RNA polymerase in bacteria: effect of extrachromosomal copies of *rrnB* on the *in vivo* localization of RNA polymerase. *J Bacteriol*, 188(11), 4007–14.
- Caramel, A. & Schnetz, K. (2000). Antagonistic control of the *Escherichia coli* bgl promoter by FIS and CAP *in vitro*. *Mol Microbiol*, 36(1), 85–92.
- Cases, I., De Lorenzo, V., & Ouzounis, C. A. (2003). Transcription regulation and environmental adaptation in bacteria. *Trends Microbiol*, 11(6), 248–253.
- Cashel, M. (1996). The stringent response. In F. C. Neidhardt (Ed.), *Escherichia coli and Salmonella* (ASM Press ed.), volume 1 (pp. 1458–1496). Washington, DC.
- Cashel, M. & Gallant, J. (1969). Two Compounds implicated in the Function of the RC Gene of *Escherichia coli*. *Nature*, 221(5183), 838–841.
- Collado-Vides, J., Magasanik, B., & Gralla, J. D. (1991). Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol Rev*, 55(3), 371–394.
- Condon, C., Philips, J., Fu, Z.-y., Squires, C., & Squires, C. L. (1992). Comparison of the expression of the RNA operons in *Escherichia coli*. 1(1), 4175–4185.
- Costanzo, A., Nicoloff, H., Barchinger, S. E., Banta, A. B., Gourse, R. L., & Ades, S. E. (2008). ppGpp and DksA likely regulate the activity of the extracytoplasmic stress factor sigmaE in *Escherichia coli* by both direct and indirect mechanisms. *Mol Microbiol*, 67(3), 619–632.
- Dame, R. T. (2005). The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol Microbiol*, 56(4), 858–870.
- de Jong, H., Ranquet, C., Ropers, D., Pinel, C., & Geiselmann, J. (2010). Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria. *BMC Syst Biol*, 4, 55.
- De Las Peñas, A., Connolly, L., & Gross, C. A. (1997). SigmaE is an essential sigma factor in *Escherichia coli*. *J Bacteriol*, 179(21), 6862–4.
- Dennis, P. P., Nene, V., & Glass, R. E. (1985). Autogenous posttranscriptional regulation of RNA polymerase beta and beta' subunit synthesis in *Escherichia coli*. *J Bacteriol*, 161(2), 803–6.
- Dong, T., Kirchhof, M. G., & Schellhorn, H. E. (2008). RpoS regulation of gene expression during exponential growth of *Escherichia coli* K12. *Mol Genet Genomics*, 279(3), 267–277.

- Dong, T. & Schellhorn, H. E. (2009). Control of RpoS in global gene expression of *Escherichia coli* in minimal media. *Mol Genet Genomics*, 281(1), 19–33.
- Dykxhoorn, D. M., Pierre, R. S., Linn, T., & St Pierre, R. (1996). Synthesis of the beta and beta' subunits of *Escherichia coli* RNA polymerase is autogenously regulated *in vivo* by both transcriptional and translational mechanisms. *Mol Microbiol*, 19(3), 483–493.
- Ellwood, M. & Nomura, M. (1982). Chromosomal locations of the genes for rRNA in *Escherichia coli* K-12. *J Bacteriol*, 149(2), 458–468.
- Engbaek, F., Gross, C., & Burgess, R. R. (1976). Biosynthesis of *Escherichia coli* RNA polymerase subunits upon release of rifampicin inhibition. *Mol Gen Genet*, 143(3), 297–9.
- Engelhorn, M. & Geiselmann, J. (1998). Maximal transcriptional activation by the IHF protein of *Escherichia coli* depends on optimal DNA bending by the activator. *Mol Microbiol*, 30(2), 431–41.
- England, P., Westblade, L. F., Karimova, G., Robbe-Saule, V., Norel, F., & Kolb, A. (2008). Binding of the unorthodox transcription activator, Crl, to the components of the transcription machinery. *J Biol Chem*, 283(48), 33455–64.
- Farewell, A., Kvint, K., & Nyström, T. (1998). Negative regulation by RpoS: a case of sigma factor competition. *Mol Microbiol*, 29(4), 1039–51.
- Finkel, S. E. & Johnson, R. C. (1993). The Fis protein: it's not just for DNA inversion anymore. *Mol Microbiol*, 6(22), 3257–3265.
- Freyre-Gonzalez, J. A. & Trevino-Quintanilla, L. G. (2010). Analyzing Regulatory Networks in Bacteria. *Nature education*, 3(9), 24.
- Fukuda, R., Taketo, M., & Ishihama, A. (1978). Autogenous regulation of RNA polymerase beta subunit synthesis *in vitro*. *J Biol Chem*, 253(13), 4501–4.
- Gaal, T., Mandel, M. J., Silhavy, T. J., & Gourse, R. L. (2006). Crl facilitates RNA polymerase holoenzyme formation. *J Bacteriol*, 188(22), 7966–7970.
- Giladi, H., Koby, S., Prag, G., Engelhorn, M., Geiselmann, J., & Oppenheim, a. B. (1998). Participation of IHF and a distant UP element in the stimulation of the phage P L promoter. *Mol Microbiol*, 30(2), 443–51.
- Gilbert, S. F., Boer, H. A. D., de Boer, H. A., & Nomura, M. (1979). Identification of initiation sites for the *in vitro* transcription of rRNA operons rrnE and rrnA in *E. coli*. *Cell*, 17(1), 211–24.

- Gilbert, W. & Müller-Hill, B. (1966). Isolation of the *lac* repressor. *Proc Natl Acad Sci U S A*, 56(6), 1891–8.
- Glaser, G. & Cashel, M. (1979). *in vitro* transcripts from the *rrnB* ribosomal RNA cistron originate from two tandem promoters. *Cell*, 16(1), 111–21.
- Goosen, N., Putte, P. V. D., & Van De Putte, P. (1995). The regulation of transcription initiation by integration host factor. *Mol Microbiol*, 16(1), 1–7.
- Gourse, R. L. (1988). Visualization and quantitative analysis of complex formation between *E. coli* RNA polymerase and an rRNA promoter *in vitro*. *Nucleic Acids Res*, 16(20), 9789–809.
- Gourse, R. L., de Boer, H. a., & Nomura, M. (1986). DNA determinants of rRNA synthesis in *E. coli*: growth rate dependent regulation, feedback inhibition, upstream activation, antitermination. *Cell*, 44(1), 197–205.
- Gourse, R. L., Ross, W., & Rutherford, S. T. (2006). General pathway for turning on promoters transcribed by RNA polymerases containing alternative sigma factors. *J Bacteriol*, 188(13), 4589–91.
- Grainger, D. C. & Busby, S. J. W. (2008). Global regulators of transcription in *Escherichia coli*: mechanisms of action and methods for study. *Adv Appl Microbiol*, 65, 93–113.
- Grainger, D. C., Hurd, D., Harrison, M., Holdstock, J., & Busby, S. J. W. (2005). Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc Natl Acad Sci U S A*, 102(49), 17693–8.
- Grigorova, I. L., Phleger, N. J., Mutalik, V. K., & Gross, C. a. (2006). Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. *Proc Natl Acad Sci U S A*, 103(14), 5332–7.
- Gross, C. A., Lonetto, M., & Losick, R. (1992). Bacterial sigma factors. In S. L. McKnight (Ed.), *Transcriptional Regulation* (Cold Spring ed.). (pp. 129–176). New York, NY.
- Hara, A. & Sy, J. (1983). Guanosine 5'-triphosphate, 3'-diphosphate 5'-phosphohydrolase. Purification and substrate specificity. *J Biol Chem*, 258(3), 1678–83.
- Haseltine, W. A., Block, R., Gilbert, W., & Weber, K. (1972). MSI and MSII made on ribosome in idling step of protein synthesis. *Nature*, 238(5364), 381–4.
- Haugen, S. P., Ross, W., & Gourse, R. L. (2008). Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat Rev Microbiol*, 6(7), 507–19.
- Hayward, R. S. & Fyfe, S. (1978). Over-synthesis and instability of sigma protein in a merodiploid strain of *Escherichia coli*. *Mol Gen Genet*, 159(1), 89–99.

- Hernandez, V. J. & Bremer, H. (1991). *Escherichia coli* ppGpp synthetase II activity requires *spoT*. *J Biol Chem*, 266(9), 5991–9.
- Hill, T. M., Sharma, B., Valjavec-Gratian, M., & Smith, J. (1997). *sfi*-independent filamentation in *Escherichia coli* Is *lexA* dependent and requires DNA damage for induction. *J Bacteriol*, 179(6), 1931–9.
- Hinshelwood, C. (1946). The chemical kinetics of the bacterial cell.
- Hirokawa, G., Nijman, R. M., Raj, V. S., Kaji, H., Igarashi, K., & Kaji, A. (2005). The role of ribosome recycling factor in dissociation of 70S ribosomes into subunits. *RNA*, 11(8), 1317–1328.
- Honda, Y., Sakai, H., Hiasa, H., Tanaka, K., Komano, T., & Bagdasarian, M. (1991). Functional division and reconstruction of a plasmid replication origin: molecular dissection of the *oriV* of the broad-host-range plasmid RSF1010. *Proc Natl Acad Sci U S A*, 88(1), 179–83.
- Ishihama, A. (1999). Modulation of the nucleoid, the transcription apparatus, and the translation machinery in bacteria for stationary phase survival. *Genes Cells*, 4(3), 135–143.
- Ishihama, A. (2000). Functional modulation of *Escherichia coli* RNA polymerase. *Annu Rev Microbiol*, 54, 499–518.
- Iwakura, Y., Ito, K., & Ishihama, A. (1974). Biosynthesis of RNA polymerase in *Escherichia coli*. I. Control of RNA polymerase content at various growth rates. *Mol Gen Genet*, 133(1), 1–23.
- Jacob, F. (1970). *La Logique du Vivant, Une Histoire de L'Hérédité*. (Gallimard ed.). Paris.
- Jensen, K. F. & Pedersen, S. (1990). Metabolic growth rate control in *Escherichia coli* may be a consequence of subsaturation of the macromolecular biosynthetic apparatus with substrates and catalytic components. *Microbiol Rev*, 54(2), 89–100.
- Jin, D. J., Cagliero, C., & Zhou, Y. N. (2011). Growth rate regulation in *Escherichia coli*. *FEMS Microbiol Rev*, 36(2), 269–87.
- Jishage, M., Iwata, a., Ueda, S., & Ishihama, a. (1996). Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of four species of sigma subunit under various growth conditions. *J Bacteriol*, 178(18), 5447–51.
- Jishage, M., Kvint, K., Shingler, V., & Nyström, T. (2002). Regulation of sigma factor competition by the alarmone ppGpp. *Genes Dev*, 16(10), 1260–1270.

- Jordi, B. J., Fielder, A. E., Burns, C. M., Hinton, J. C., Dover, N., Ussery, D. W., & Higgins, C. F. (1997). DNA binding is not sufficient for H-NS-mediated repression of proU expression. *J Biol Chem*, 272(18), 12083–12090.
- Justice, S. S., Hunstad, D. A., Seed, P. C., & Hultgren, S. J. (2006). Filamentation by *Escherichia coli* subverts innate defenses during urinary tract infection. *Proc Natl Acad Sci U S A*, 103(52), 19884–9.
- Kajitani, M., Fukuda, R., & Ishihama, A. (1980). Autogenous and post-transcriptional regulation of *Escherichia coli* RNA polymerase synthesis *in vitro*. *Mol Gen Genet*, 179(3), 489–96.
- Keener, J. & Nomura, M. (1996). Regulation of ribosome synthesis. In F. C. Neidhardt (Ed.), *Escherichia coli and Salmonella* (pp. 1417–1431).
- Klauck, E., Typas, A., & Hengge, R. (2007). The sigmaS subunit of RNA polymerase as a signal integrator and network master regulator in the general stress response in *Escherichia coli*. *Sci Prog*, 90(Pt 2-3), 103–27.
- Klumpp, S., Zhang, Z., & Hwa, T. (2009). Growth rate-dependent global effects on gene expression in bacteria. *Cell*, 139(7), 1366–1375.
- Koga, K., Harada, T., Shimizu, H., & Tanaka, K. (2005). Bacterial luciferase activity and the intracellular redox pool in *Escherichia coli*. *Mol Genet Genomics*, 274(2), 180–8.
- Lacour, S. & Landini, P. (2004). SigmaS-dependent gene expression at the onset of stationary phase in *Escherichia coli*: function of sigmaS-dependent genes and identification of their promoter sequences. *J Bacteriol*, 186(21), 7186–95.
- Lange, R. & Hengge-Aronis, R. (1994). The cellular concentration of the sigma S subunit of RNA polymerase in *Escherichia coli* is controlled at the levels of transcription, translation, and protein stability. *Genes Dev*, 8(13), 1600–1612.
- Laurie, A. D., Bernardo, L. M. D., Sze, C. C., Skarfstad, E., Szalewska-Palasz, A., Nyström, T., & Shingler, V. (2003). The role of the alarmone (p)ppGpp in sigma N competition for core RNA polymerase. *J Biol Chem*, 278(3), 1494–1503.
- Lelong, C., Aguiluz, K., Luche, S., Kuhn, L., Garin, J., Rabilloud, T., & Geiselmann, J. (2007). The Crl-RpoS regulon of *Escherichia coli*. *Mol Cell Proteomics*, 6(4), 648–659.
- Lewis, D. E., Geanakopulos, M., & Adhya, S. (1999). Role of HU and DNA supercoiling in transcription repression: specialized nucleoprotein repression complex at *gal* promoters in *Escherichia coli*. *Mol Microbiol*, 31(2), 451–461.
- Liang, S., Bipatnath, M., Xu, Y., Chen, S., Dennis, P., Ehrenberg, M., & Bremer, H. (1999). Activities of constitutive promoters in *Escherichia coli*. *J Mol Biol*, 292(1), 19–37.

- Lund, E. & Dahlberg, J. E. (1979). Initiation of *Escherichia coli* ribosomal RNA synthesis *in vivo*. *Proc Natl Acad Sci U S A*, 76(11), 5480–5484.
- Maaloe, O. (1979). Regulation of the protein-synthesizing machinery - ribosomes, tRNA, factors, and so on. In R. Goldberger (Ed.), *Biological Regulation and Development* (pp. 487–542). New York: Plenum Press.
- Madan Babu, M. (2003). Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res*, 31(4), 1234–1244.
- Madigan, M. T. & Martinko, J. M. (2006). *Brock Biology of Microorganisms*. Pearson Prentice Hall.
- Maeda, H., Fujita, N., & Ishihama, A. (2000). Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res*, 28(18), 3497–503.
- Martínez-Antonio, A. & Collado-Vides, J. (2003). Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol*, 6(5), 482–489.
- Mascher, T., Helmann, J. D., & Uden, G. (2006). Stimulus Perception in Bacterial Signal-Transducing Histidine Kinases. *Microbiol Mol Biol Rev*, 70(4), 910–938.
- McLeod, S. M. & Johnson, R. C. (2001). Control of transcription by nucleoid proteins. *Curr Opin Microbiol*, 4(2), 152–9.
- Meek, D. W. & Hayward, R. S. (1986). Direct evidence for autogenous regulation of the *Escherichia coli* genes *rpoBC* *in vivo*. *Mol Gen Genet*, 202(3), 500–8.
- Molina, N. & Van Nimwegen, E. (2008). Universal patterns of purifying selection at noncoding positions in bacteria. *Genome Res*, 18(1), 148–160.
- Monod, J. (1949). The growth of bacterial cultures. (x).
- Murray, K. D. & Bremer, H. (1996). Control of *spoT*-dependent ppGpp synthesis and degradation in *Escherichia coli*. *J Mol Biol*, 259(1), 41–57.
- Nomura, M., Gourse, R., & Baughman, G. (1984). Regulation of the synthesis of ribosomes and ribosomal components. *Annu Rev Biochem*, 53, 75–117.
- Novick, A. & Szilard, L. (1950). Description of the chemostat. *Science*, 112(2920), 715–6.
- Nyström, T. (2004). Growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition? *Mol Microbiol*, 54(4), 855–862.
- Passador, L. & Linn, T. (1989). Autogenous regulation of the RNA polymerase beta subunit of *Escherichia coli* occurs at the translational level *in vivo*. *J Bacteriol*, 171(11), 6234–42.

- Patten, C. L., Kirchhof, M. G., Schertzberg, M. R., Morton, R. A., & Schellhorn, H. E. (2004). Microarray analysis of RpoS-mediated gene expression in *Escherichia coli* K-12. *Mol Genet Genomics*, 272(5), 580–91.
- Paul, B. J., Barker, M. M., Ross, W., Schneider, D. a., Webb, C., Foster, J. W., & Gourse, R. L. (2004). DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell*, 118(3), 311–22.
- Paul, B. J., Berkmen, M. B., & Gourse, R. L. (2005). DksA potentiates direct activation of amino acid promoters by ppGpp. *Proc Natl Acad Sci U S A*, 102(22), 7823–8.
- Paul, B. J., Ross, W., Gaal, T., & Gourse, R. L. (2004). rRNA transcription in *Escherichia coli*. *Annu Rev Genet*, 38, 749–70.
- Peacock, S., Cenatiempo, Y., Robakis, N., Brot, N., & Weissbach, H. (1982). *in vitro* synthesis of the first dipeptide of the beta subunit of *Escherichia coli* RNA polymerase. *Proc Natl Acad Sci U S A*, 79(15), 4609–12.
- Pedersen, S. & Reeh, S. V. (1976). Analysis of the proteins synthesized in ultraviolet light-irradiated *Escherichia coli* following infection with the bacteriophages lambdadrifd 18 and lambdadfus-3. *Mol Gen Genet*, 144(3), 339–43.
- Pérez-Rueda, E. & Collado-Vides, J. (2000). The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res*, 28(8), 1838–47.
- Pettijohn, D. E. (1996). The Nucleoid. In F. Neidhardt (Ed.), *Escherichia coli and Salmonella : cellular and molecular biology* (pp. 158–166).
- Piper, S. E., Mitchell, J. E., Lee, D. J., & Busby, S. J. W. (2009). A global view of *Escherichia coli* Rsd protein and its interactions. *Mol Biosyst*, 5(12), 1943–7.
- Potrykus, K. & Cashel, M. (2008). (p)ppGpp: still magical? *Annu Rev Microbiol*, 62, 35–51.
- Pratt, L. A. & Silhavy, T. J. (1998). Crl stimulates RpoS activity during stationary phase. *Mol Microbiol*, 29(5), 1225–1236.
- Ptashne, M. & Gaan, A. (2002). *Genes and Signals*. Cold Spring Harbor, NY.
- Record, M. T., Reznikoff, W. S., Craig, M. L., Mcquade, K. L., & Schlax, P. J. (1996). *Escherichia coli* RNA Polymerase (Es70), Promoters, and the Kinetics of the Steps of Transcription Initiation. In *Escherichia coli and Salmonella* (ASM Press ed.), volume 2 chapter 54, (pp. 792–820). Washington, D.C: Neidhardt, F C Curtiss III, R Ingraham, J L Lin, E C C Low, K B Magasanik, B Reznikoff, W S Riley, M Schaechter, M Umbarger, H E.

- Robbe-Saule, V., Carreira, I., Kolb, A., & Norel, F. (2008). Effect of growth temperature on Crl-dependent regulation of sigmaS activity in *Salmonella enterica* serovar Typhimurium. *J Bacteriol*, 190(13), 4453–4459.
- Robbe-Saule, V., Jaumouillé, V., Prévost, M.-C., Guadagnini, S., Talhouarne, C., Mathout, H., Kolb, A., & Norel, F. (2006). Crl activates transcription initiation of RpoS-regulated genes involved in the multicellular behavior of *Salmonella enterica* serovar Typhimurium. *J Bacteriol*, 188(11), 3983–3994.
- Robbe-Saule, V., Lopes, M. D., Kolb, A., & Norel, F. (2007). Physiological effects of Crl in *Salmonella* are modulated by sigmaS level and promoter specificity. *J Bacteriol*, 189(8), 2976–87.
- Ryals, J., Little, R., & Bremer, H. (1982). Control of rRNA and tRNA syntheses in *Escherichia coli* by guanosine tetraphosphate. *J Bacteriol*, 151(3), 1261–8.
- Salgado, H., Santos-Zavaleta, A., Gama-Castro, S., Peralta-Gil, M., Peñaloza Spínola, M. I., Martínez-Antonio, A., Karp, P. D., & Collado-Vides, J. (2006). The comprehensive updated regulatory network of *Escherichia coli* K-12. *BMC bioinformatics*, 7, 5.
- Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press.
- Sarmientos, P. & Cashel, M. (1983). Carbon starvation and growth rate-dependent regulation of the *Escherichia coli* ribosomal RNA promoters: differential control of dual promoters. *Proc Natl Acad Sci U S A*, 80(22), 7010–3.
- Schaechter, M., Maaloe, O., & Kjeldgaard, N. O. (1958). Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of *Salmonella typhimurium*. *J Gen Microbiol*, 19(3), 592–606.
- Schneider, R., Travers, A., Kutateladze, T., & Muskhelishvili, G. (1999). A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. *Mol Microbiol*, 34(5), 953–964.
- Schröder, O. & Wagner, R. (2000). The bacterial DNA-binding protein H-NS represses ribosomal RNA transcription by trapping RNA polymerase in the initiation complex. *J Mol Biol*, 298(5), 737–48.
- Sokolov, A., Apodaca, M. M., Grzybowski, B. a., & Aranson, I. S. (2010). Swimming bacteria power microscopic gears. *Proc Natl Acad Sci U S A*, 107(3), 969–74.
- Srivatsan, A. & Wang, J. D. (2008). Control of bacterial transcription, translation and replication by (p)ppGpp. *Curr Opin Microbiol*, 11(2), 100–5.

- Steinhaus, E. A. & Birkeland, J. M. (1939). Studies on the Life and Death of Bacteria: I. The Senescent Phase in Aging Cultures and the Probable Mechanisms Involved. *J Bacteriol*, 38(3), 249–61.
- Stent, G. S. & Benner, S. (1961). A genetic locus for the regulation of ribonucleic acid synthesis. *Proc Natl Acad Sci U S A*, 47, 2005–14.
- Steward, K. L. & Linn, T. (1991). *in vivo* analysis of overlapping transcription units in the *rplKAJLrpoBC* ribosomal protein-RNA polymerase gene cluster of *Escherichia coli*. *J Mol Biol*, 218(1), 23–31.
- Steward, K. L. & Linn, T. (1992). Transcription frequency modulates the efficiency of an attenuator preceding the *rpoBC* RNA polymerase genes of *Escherichia coli*: possible autogenous control. *Nucleic Acids Res*, 20(18), 4773–9.
- Tissieres, A. & Watson, J. D. (1958). Ribonucleoprotein particles from *Escherichia coli*. *Nature*, 182(4638), 778–780.
- Typas, A., Barembruch, C., Possling, A., & Hengge, R. (2007). Stationary phase reorganisation of the *Escherichia coli* transcription machinery by Crl protein, a fine-tuner of sigmas activity and levels. *EMBO J*, 26(6), 1569–1578.
- Typas, A., Stella, S., Johnson, R. C., & Hengge, R. (2007). The -35 sequence location and the Fis-sigma factor interface determine sigmas selectivity of the proP (P2) promoter in *Escherichia coli*. *Mol Microbiol*, 63(3), 780–96.
- Ueta, M., Yoshida, H., Wada, C., Baba, T., Mori, H., & Wada, A. (2005). Ribosome binding proteins YhbH and YfiA have opposite functions during 100S formation in the stationary phase of *Escherichia coli*. *Genes Cells*, 10(12), 1103–1112.
- Ulrich, L. E., Koonin, E. V., & Zhulin, I. B. (2005). One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol*, 13(2), 52–56.
- Volkmer, B. & Heinemann, M. (2011). Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. *PLoS One*, 6(7), e23126.
- Wada, A., Igarashi, K., Yoshimura, S., Aimoto, S., & Ishihama, A. (1995). Ribosome modulation factor: stationary growth phase-specific inhibitor of ribosome functions from *Escherichia coli*. *Biochem Biophys Res Commun*, 214(2), 410–417.
- Wang, P., Robert, L., Pelletier, J., Dang, W. L., Taddei, F., Wright, A., & Jun, S. (2010). Robust growth of *Escherichia coli*. *Curr Biol*, 20(12), 1099–103.
- Wanner, B. L., Kodaira, R., & Neidhardt, F. C. (1977). Physiological regulation of a decontrolled lac operon. *J Bacteriol*, 130(1), 212–222.

- Wassarman, K. M. (2007). 6S RNA: a regulator of transcription. *Mol Microbiol*, 65(6), 1425–31.
- Weber, H., Polen, T., Heuveling, J., Wendisch, V. F., & Hengge, R. (2005). Genome-wide analysis of the general stress response network in *Escherichia coli*: sigmaS-dependent genes, promoters, and sigma factor selectivity. *J Bacteriol*, 187(5), 1591–603.
- Xu, J. & Johnson, R. C. (1995). Fis activates the RpoS-dependent stationary-phase expression of *proP* in *Escherichia coli*. *J Bacteriol*, 177(18), 5222–5231.
- Yamagishi, M., Matsushima, H., Wada, A., Sakagami, M., Fujita, N., & Ishihama, A. (1993). Regulation of the *Escherichia coli* *rmf* gene encoding the ribosome modulation factor: growth phase- and growth rate-dependent control. *EMBO J*, 12(2), 625–630.
- Yoshida, H., Maki, Y., Kato, H., Fujisawa, H., Izutsu, K., Wada, C., & Wada, A. (2002). The ribosome modulation factor (RMF) binding site on the 100S ribosome of *Escherichia coli*. *J Biochem*, 132(6), 983–989.
- Yoshida, H., Yamamoto, H., Uchiumi, T., & Wada, A. (2004). RMF inactivates ribosomes by covering the peptidyl transferase centre and entrance of peptide exit tunnel. *Genes Cells*, 9(4), 271–278.
- Yura, T. & Ishihama, A. (1979). Genetics of bacterial RNA polymerases. *Annu Rev Genet*, 13, 59–97.
- Zambrano, M. M., Siegele, D. A., Almirón, M., Tormo, A., & Kolter, R. (1993). Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. *Science*, 259(5102), 1757–60.
- Zhou, Y. N. & Gross, C. A. (1992). How a mutation in the gene encoding sigma 70 suppresses the defective heat shock response caused by a mutation in the gene encoding sigma 32. *J Bacteriol*, 174(22), 7128–37.

Résumé

La faculté d'adaptation aux conditions environnementales des bactéries provient de la complexité de leur réseau de régulation génique, impliquant de nombreux régulateurs spécifiques et la machinerie d'expression génique. Nous avons montré que le gène *crl*, codant un régulateur global d'*Escherichia coli*, est exprimé de façon transitoire lors de la phase exponentielle. Notre étude a permis d'identifier deux régulateurs responsables de ce profil parmi une centaine testés. Ainsi, le complexe CRP-AMPc, réprime de façon indirecte l'expression de *crl*, tandis que la nucléoprotéine Fis se fixe sur le promoteur de *crl* et active sa transcription directement. Le profil d'expression de *crl* étant similaire à celui de nombreux régulateurs globaux, nous nous sommes intéressés au rôle de la machinerie globale d'expression des gènes et à son impact sur la croissance. Dans ce but, nous avons construit un système nous permettant de contrôler la croissance d'*E. coli* en modulant l'expression des sous-unités β et β' de l'ARN polymérase et donc le niveau de transcription dans la cellule. Lorsque l'ARN polymérase est en faible concentration, le taux de croissance devient quasiment nul et les cellules filamentent. Ce contrôle de la croissance est dose dépendant et a été mis en évidence autant à l'échelle de la population qu'à celle de la cellule unique. Nous avons enfin étudié par RNA-seq l'impact du niveau d'ARN polymérase sur la transcription de l'ensemble du génome de cette souche. Cette étude montre que toutes les classes fonctionnelles de gènes sont affectées par notre système, à l'exception des gènes qui codent les protéines ribosomales.

Summary

Bacteria can adapt to many different environmental conditions. This capacity of adaptation is conferred to the organism by a complex regulatory network, composed of specific regulators and the global gene expression machinery. We have studied the expression dynamics of Crl, a global regulator of *Escherichia coli*, and observed a peak of transcription during the exponential phase of growth. In order to identify potential regulators of *crl* expression, we have measured the expression profile of *crl* in about one hundred different mutant strains. This screen has revealed that CRP-cAMP represses indirectly the transcription of *crl* and the nucleoprotein Fis activates transcription of the *crl* promoter by binding to the *crl* promoter region. We noted that the expression of most global regulators of *E. coli* have an expression profile similar to the one of Crl. We have therefore studied the relationship between global gene expression machinery and cellular growth. We constructed a bacterium where the transcription of the two large subunits of RNA polymerase, β et β' , is under external control. A small concentration of RNA polymerase leads to a small growth rate of this engineered bacterium and the cells start to filament, whereas a high concentration of RNA polymerase produces phenotypically wild-type cells. We have characterized the control of growth rate by our system at the population level and in single cells. An analysis of the global transcription pattern of this strain by RNA-seq shows that the transcription of genes in all functional classes, with the possible exception of genes coding for ribosomal proteins, are almost equally affected by the modifications of the intracellular concentration of RNA polymerase.